Asymptotic normality of the Quasi Maximum Likelihood Estimator for multidimensional causal processes

Jean-Marc Bardet, Olivier Wintenberger

Submitted to the Annals of Statistics
arXiv: math.PR/0000000

ASYMPTOTIC NORMALITY OF THE QUASI MAXIMUM LIKELIHOOD ESTIMATOR FOR MULTIDIMENSIONAL CAUSAL PROCESSES

By Jean-Marc Bardet and Olivier Wintenberger

CES (SAMOS-Matisse), University Paris 1 Panthéon-Sorbonne, France

Strong consistency and asymptotic normality of the Quasi-Maximum Likelihood Estimator (QMLE) are given for a general class of multidimensional causal processes. For particular cases already studied in the literature (for instance univariate or multivariate GARCH, ARCH, ARMA-GARCH processes) the assumptions required for establishing these results are often weaker than existing conditions. The QMLE asymptotic behavior is also given for numerous new examples of univariate or multivariate processes (for instance TARCH or NLARCH processes).

1. Introduction. In this paper the asymptotic behavior of the Quasi-Maximum Likelihood Estimator (QMLE) is studied for general $\mathbb{R}^m$-valued stationary process. The time series $X = (X_t,\ t \in \mathbb{Z})$ is defined as a solution of the equation:

$$X_t = M_{\theta_0}(X_{t-1}, X_{t-2}, \ldots) \cdot \xi_t + f_{\theta_0}(X_{t-1}, X_{t-2}, \ldots), \quad \forall\, t \in \mathbb{Z}, \tag{1.1}$$

almost everywhere (a.e.). Here $M_{\theta_0}(X_{t-1}, X_{t-2}, \ldots)$ is a $(m \times p)$-random matrix having almost surely (a.s.) full rank m, the sequence $(\xi_t)_{t\in\mathbb{Z}}$ of $\mathbb{R}^p$-random vectors $(\xi_t^{(k)})_{1\le k\le p}$ are independent and identically distributed satisfying $\mathbb{E}\big[\xi_0^{(k)}\xi_0^{(k')}\big] = 0$ for $k \neq k'$ and $\mathbb{E}\big[{\xi_0^{(k)}}^2\big] = \mathrm{Var}\,(\xi_0^{(k)}) = 1$ and $f_{\theta_0}(X_{t-1}, X_{t-2}, \ldots)$ is a sequence of $\mathbb{R}^m$-random vectors. Various popular econometric time series models can be written in the form (1.1). The case $f_\theta \equiv 0$ and

$$\begin{aligned} H_\theta(X_{t-1}, X_{t-2}, \ldots) := C_0 &+ \sum_{i=1}^{q'}\sum_{j=1}^{k} C_{ij} X_{t-i} X'_{t-i} C'_{ij} \\ &+ \sum_{i=1}^{q}\sum_{j=1}^{k} D_{ij} H_\theta(X_{t-i-1}, X_{t-i-2}, \ldots) D'_{ij}, \ \text{where}^1 H_\theta := M_\theta \cdot M'_\theta, \end{aligned} \tag{1.2}$$

AMS 2000 subject classifications: Primary 62M10, 62F12
Keywords and phrases: Quasi-Maximum Likelihood Estimator; Strong consistency; Asymptotic normality; Multidimensional causal processes; Multivariate GARCH processes

corresponds to the BEKK representation of multivariate GARCH(q,q') defined by Engle and Kroner [13], see also Bollerslev [4]. Their natural generalization,

$$H_\theta(X_{t-1},X_{t-2},\ldots):=B_0+\sum_{i=1}^{\infty}B_iX_{t-i}X_{t-i}'B_i',$$

defines the multivariate ARCH(∞) processes. If $M_\theta\equiv I_d$, a process X satisfying relation (1.1) is a multivariate Non Linear AR(∞) process.

Various methods can be employed to estimate the unknown parameter θ_0. Maximum Likelihood Estimation (MLE) is a common one. Several authors studied the asymptotic behavior of MLE for particular cases of multivariate processes satisfying (1.1), see for instance Bollerslev and Wooldridge [5], Jeantheau [19] for multivariate GARCH(q,q') processes and Dunsmuir and Hannan [11], Mauricio [22] for multivariate ARMA processes. A proof of the efficiency of those estimators was obtained in Berkes and Horváth [1], in the case of one-dimensional GARCH(q,q'). Even if the convergence rate of the MLE can be optimal this method presents numerous drawbacks. For example, the conditional likelihood depends on the distribution of the innovations ξ_t, which is often unknown, and on all the past values of the process X, which are unobserved.

In the present paper we consider an approximation of the MLE called Quasi-Maximum Likelihood Estimation (QMLE). If the sequence $(\xi_t)_{t\in\mathbb{Z}}$ is a sequence of standardized Gaussian vectors, the conditional likelihood of X is, up to an additional constant, equal to

$$(1.3)\quad L_n(\theta):=-\frac{1}{2}\sum_{t=1}^{n}q_t(\theta)\qquad\text{for all }\theta\in\Theta$$
$$\text{with }\ q_t(\theta):=\Big[(X_t-f_\theta^t)'\big(H_\theta^t\big)^{-1}(X_t-f_\theta^t)+\ \log\big(\det\big(H_\theta^t\big)\big)\Big],$$

$f_\theta^t=f_\theta(X_{t-1},X_{t-2},\ldots)$, $M_\theta^t=M_\theta(X_{t-1},X_{t-2},\ldots)$ and $H_\theta^t:=M_\theta^tM_\theta^{t\prime}$.

¿From now on we omit any assumption on the distribution of the ξ_t. The QMLE is obtained by plugging in the likelihood the approximations $\widehat{f}_\theta^t:=f_\theta(X_{t-1},\ldots,X_1,u)$, $\widehat{M}_\theta^t:=M_\theta(X_{t-1},\ldots,X_1,u)$ and $\widehat{H}_\theta^t:=\widehat{M}_\theta^t\cdot(\widehat{M}_\theta^t)'$

[1] Here A' is the transpose of the matrix A.

where u is a finitely-non-zero sequence[2] $(u_n)_{n\in\mathbb{N}}$:

$$\text{(1.4)}\quad \widehat{L}_n(\theta) := -\frac{1}{2}\sum_{t=1}^{n}\widehat{q}_t(\theta)$$
$$\text{with}\quad \widehat{q}_t(\theta) := \Big[\big(X_t-\widehat{f}^t_\theta\big)'\big(\widehat{H}^t_\theta\big)^{-1}\big(X_t-\widehat{f}^t_\theta\big)+\log\big(\det\big(\widehat{H}^t_\theta\big)\big)\Big].$$

The QMLE $\widehat{\theta}_n$ is the M-estimator associated with the quasi-likelihood $\widehat{L}_n$ given as the maximizer

$$\widehat{\theta}_n := \operatorname*{Argmax}_{\theta\in\Theta}\ \widehat{L}_n(\theta). \tag{1.5}$$

A basic idea of this paper is to restrict the set of parameters Θ in such a way that moment conditions on ξ_0 imply both the existence of a solution X and finite moments of sufficiently high order for X. This strategy is available for the very general model (1.1) thanks to a result of Doukhan and Wintenberger [10], see Section 2. Then we use the moment conditions to settle both consistency and asymptotic normality, see Section 3.

We restrict the set of the parameters in such a way that we only assume finite moments of orders 2 or 4 on ξ_0, which are necessary conditions for consistency or asymptotic normality, respectively, see for example Straumann and Mikosch [26] for some particular classes of non-linear time series models. In turn, these conditions guarantee the existence of moments of order 2 or 4 of X, respectively. Notice that for one-dimensional GARCH models these moment conditions on X can be relaxed, see Francq and Zakoïan [15], Berkes *et al.* [2]. For Markovian models, Straumann and Mikosch [26] achieved the asymptotic normality assuming moment conditions but the corresponding restriction on Θ is non-explicit except for the AGARCH models. In the case of ARCH(∞), the conditions are not comparable with those in Robinson and Zaffaroni [25]. Our restriction on Θ is stronger whereas we sharpen the moment conditions of order $2+\delta$ to the order 2 on ξ_0 for the strong consistency. Finally, for multivariate models the conditions are sharper than those in Comte and Lieberman [7] and Ling and McAleer [21] who derived the asymptotic normality for particular models under moments of order 4, 6 or 8 on X. In Section 4 we provide for the first time the consistency and asymptotic normality of the QMLE in TARCH, NLARCH and Non Linear AR(∞) models.

But to begin with, the following Section 2 deals with the various assumptions on the general model (1.1) that are needed.

[2]This means that $u_n \neq 0$ only for finitely many $n \in \mathbb{N}$.

2. Notation and assumptions. In the sequel, some standard notation is used:

- The symbol $\|.\|$ denotes the usual Euclidean norm of a vector or a matrix (for A a $(n\times p)$-matrix, $\|A\|=\sup_{\|Y\|\leq 1}\{\|AY\|,\ Y\in\mathbb{R}^p\}$);
- For the measurable vector- or matrix-valued function g defined on Θ, $\|g\|_\Theta=\sup_{\theta\in\Theta}\|g(\theta)\|$;
- If V is a vector space then V^∞ denotes the set of the finitely-non-zero sequences x i.e., there exists $N>0$ such that $x=(x_1,x_2,\ldots,x_N,0,0,\ldots)$;
- The symbol 0 denotes the null sequence in $\mathbb{R}^{\mathbb{N}}$;
- If V is a Banach space and Θ is a subset of $\mathbb{R}^d$ then $\mathcal{C}(\Theta,V)$ denotes the Banach space of V-valued continuous functions on Θ equipped with the uniform norm $\|\cdot\|_\Theta$ and $\mathbb{L}^r(\mathcal{C}(\Theta,V))$ $(r\geq 1)$ denotes the Banach space of random a.e. continuous functions f such that $\mathbb{E}\big[\|f\|_\theta^r\big]<\infty$.

2.1. *Definition of the parameter sets $\Theta(r)$ and $\widetilde{\Theta}(r)$.* In proposition 1 below we provide the existence of a stationary solution of the general model (1.1). Two conditions of different types are used: the first one is a Lipschitz condition on the functions f and M in (1.1), the second one is a restriction on the set of the parameters.

Let us assume that for any $\theta\in\mathbb{R}^d$, $x\mapsto f_\theta(x)$ and $x\mapsto M_\theta(x)$ are Borel functions on $(\mathbb{R}^m)^\infty$ and that Rank $M_\theta(x)=m$ for all $x\in(\mathbb{R}^m)^\infty$. Assume that there exist two sequences $(\alpha_j(f,\theta))_{j\geq 1}$ and $(\alpha_j(M,\theta))_{j\geq 1}$ satisfying, for all x, y in $(\mathbb{R}^m)^\infty$,

$$\left\{\begin{array}{lcl} \|f_\theta(x)-f_\theta(y)\| & \leq & \sum_{j=1}^\infty \alpha_j(f,\theta)\|x_j-y_j\|, \\ \|M_\theta(x)-M_\theta(y)\| & \leq & \sum_{j=1}^\infty \alpha_j(M,\theta)\|x_j-y_j\|. \end{array}\right.$$

For some models, as mentioned in remark 2.1, it can be more efficient to replace the condition on M by the existence of a sequence $(\alpha_j(H,\theta))_{j\geq 1}$ such that

$$\|H_\theta(x)-H_\theta(y)\|\leq\sum_{j=1}^\infty \alpha_j(H,\theta)\|x_jx_j'-y_jy_j'\|,$$

where $H_\theta:=M_\theta\cdot M_\theta'$. Assuming $\mathbb{E}\|\xi_0\|^r<+\infty$ for some $r>0$, we can define the set

$$\Theta(r)=\left\{\theta\in\mathbb{R}^d\ \Big/\ \sum_{j=1}^\infty \alpha_j(f,\theta)+\left(\mathbb{E}\|\xi_0\|^r\right)^{1/r}\sum_{j=1}^\infty \alpha_j(M,\theta)<1\right\}. \tag{2.1}$$

This set depends on the distribution of ξ_0 via the moments $\mathbb{E}\|\xi_0\|^r$. But thanks to the fact that $\mathbb{E}\big[\xi_0^{(k)}\xi_0^{(k')}\big]=0$ for $k\neq k'$ and $\mathbb{E}\big[{\xi_0^{(k)}}^2\big]=\mathrm{Var}\,(\xi_0^{(k)})=$

1 the set $\Theta(2)$ simplifies:

$$\Theta(2)=\left\{\theta\in\mathbb{R}^d \Big/ \sum_{j=1}^{\infty}\alpha_j(f,\theta)+\sqrt{p}\sum_{j=1}^{\infty}\alpha_j(M,\theta)<1\right\}.$$

Proposition 1 *If $\theta_0\in\Theta(r)$ for some $r\geq 1$ there exists a unique causal (X_t is independent of $(\xi_i)_{i>t}$ for $t\in\mathbb{Z}$) solution X to the equation* (1.1) *which is stationary and ergodic and satisfies $\mathbb{E}\|X_0\|^r<\infty$.*

This result generalizes the one proved by Giraitis *et al.* [16] for ARCH(∞) models. It automatically yields weak dependence properties, see [10] for details. For such non Markovian models, the classical Lyapunov condition of Bougerol [6] cannot be applied.

Let us now consider the special cases of (1.1) where $f\equiv 0$, $m=p=1$ and there exists a Borel function $\widetilde{H}_\theta$ such that $H_\theta(x)=\widetilde{H}_\theta(x^2)$ for all $x\in\mathbb{R}^\infty$.

Corollary 1 *The result of Proposition 1 holds if $\theta_0\in\widetilde{\Theta}(r)$ for $r\geq 2$ where*

$$\widetilde{\Theta}(r)=\left\{\theta\in\mathbb{R}^d \Big/ \mathbb{E}|\xi_0|^r\Big(\sum_{j=1}^{\infty}\alpha_j(H,\theta)\Big)^{r/2}<1\right\}. \tag{2.2}$$

Remark 2.1 The ARCH(∞) process was defined by Robinson [24] as solution of the model:

$$X_t=\sigma_t\xi_t,\qquad \sigma_t^2=b_0(\theta_0)+\sum_{j=1}^{\infty}b_j(\theta_0)X_{t-j}^2, \tag{2.3}$$

where, for all $\theta\in\mathbb{R}^d$, $(b_j(\theta))_{j\geq 1}$ are sequences of non-negative real numbers. Here, $f\equiv 0$, $p=m=1$, $\alpha_j(M,\theta)=\sqrt{b_j(\theta)}$ and $\alpha_j(H,\theta)=b_j(\theta)$. Working with the set $\widetilde{\Theta}(r)$, larger than $\Theta(r)$, gives more general results.

2.2. *Uniform assumptions on Θ.* Fix some compact subset Θ of $\mathbb{R}^d$. For any sequences x, y of $(\mathbb{R}^m)^\infty$, the functions $\theta\mapsto f_\theta(x)$ and $\theta\mapsto M_\theta(x)$ are assumed to be continuous on Θ. As in [26], uniform continuity conditions on Θ are required to apply the QMLE procedure, see Lemma 1 of the Section 3. Assume that $\|f_\theta(0)\|_\Theta<\infty$ and $\|M_\theta(0)\|_\Theta<\infty$. To settle the assumptions in a short way, let us introduce the generic symbol Ψ for any of the functions f, M or H.

(A1(Ψ)) Let $\alpha_j(\Psi) = \sup_{\theta\in\Theta} \alpha_j(\Psi,\theta)$ be such that $\sum_{j\geq 1} \alpha_j(\Psi) < \infty$.

(A2) There exists $\underline{H} > 0$ such that $\inf_{\theta\in\Theta} \det\big(H_\theta(x)\big) \geq \underline{H}$ for all $x \in (\mathbb{R}^m)^\infty$.

(A3(Ψ)) The function $\theta \in \Theta \mapsto \Psi_\theta(x)$ is 2 times continuously differentiable for all $x \in (\mathbb{R}^m)^\infty$ and

$$\Big\|\frac{\partial \Psi_\theta(0)}{\partial\theta}\Big\|_\Theta + \Big\|\frac{\partial^2 \Psi_\theta(0)}{\partial\theta\partial\theta'}\Big\|_\Theta < \infty.$$

Moreover assume that there exist two integrable sequences $\big(\alpha_j^{(i)}(\Psi)\big)_{j\geq 1}$, $i = 1, 2$, such that for all x, $y \in (\mathbb{R}^m)^\infty$

$$\begin{aligned} \Big\|\frac{\partial \Psi_\theta(x)}{\partial\theta} - \frac{\partial \Psi_\theta(y)}{\partial\theta}\Big\|_\Theta &\leq \sum_{j=1}^\infty \alpha_j^{(1)}(\Psi)\|x_j - y_j\|, \\ \Big\|\frac{\partial^2 \Psi_\theta(x)}{\partial\theta\partial\theta'} - \frac{\partial^2 \Psi_\theta(y)}{\partial\theta\partial\theta'}\Big\|_\Theta &\leq \sum_{j=1}^\infty \alpha_j^{(2)}(\Psi)\|x_j - y_j\|. \end{aligned}$$

If $\Psi = H$, $\|x_j - y_j\|$ in the RHS terms is replaced with $\|x_j x_j' - y_j y_j'\|$.

The last assumption on the derivatives is just needed for the asymptotic normality of the QMLE.

2.3. *Identifiability and variance conditions.* We assume the same identifiability condition as in Jeantheau [19]:

(Id) For all $\theta \in \Theta$, ($f_\theta^t = f_{\theta_0}^t$ and $H_\theta^t = H_{\theta_0}^t$ a.s.) $\Rightarrow \theta = \theta_0$.

(Var) One of the families $(\partial f_{\theta_0}^t/\partial\theta_i)_{1\leq i\leq d}$ or $(\partial H_{\theta_0}^t/\partial\theta_i)_{1\leq i\leq d}$ is a.e. linearly independent, where:

$$\frac{\partial f_\theta^t}{\partial\theta} := \frac{\partial f_\theta}{\partial\theta}(X_{t-1},\ldots) \text{ and } \frac{\partial H_\theta^t}{\partial\theta} := \frac{\partial H_\theta}{\partial\theta}(X_{t-1},\ldots).$$

The condition **(Var)** is needed for ensuring finiteness of the asymptotic variance in the result on asymptotic normality. For ARCH(∞), Robinson and Zaffaroni [25] give sufficient assumptions for both **(Id)** and **(Var)**. They are easier to verify than **(Id)** and **(Var)** but are not as general. Alternative conditions similar to those for ARCH(∞) are not straightforward in the general model (1.1) because of its non-linear character.

3. Asymptotic behavior of the QMLE. If the model satisfies the conditions of Corollary 1, the set $\Theta(r)$ can be replaced with $\widetilde{\Theta}(r)$ in all the results of this section.

3.1. *Invertibility.* Here we follow the presentation of Straumann and Mikosch [26]. The approach of the QMLE is based on an approximation of $f_\theta^t = \mathbb{E}(X_t \mid X_{t-1}, X_{t-2}, \dots)$ and $H_\theta^t = \mathbb{E}((X_t - f_\theta^t)(X_t - f_\theta^t)' \mid X_{t-1}, X_{t-2}, \dots)$ by $\widehat{f}_\theta^t$ and $\widehat{H}_\theta^t$, defined as in the introduction. Invertibility is the property that $\widehat{f}_\theta^t$ and $\widehat{H}_\theta^t$ converge to the unobservable f_θ^t and H_θ^t, see Section 3.2 of [26] for more details. The following lemma states this result which is a necessary step in the proof of the QMLE consistency.

Lemma 1 *Assume that $\theta_0 \in \Theta(r)$ for $r \geq 2$ and that X is the stationary solution of the equation* (1.1)*.*

1. *If* **(A1(f))** *holds then $f_\theta^t \in \mathbb{L}^r(\mathcal{C}(\Theta, \mathbb{R}^m))$ and*

$$\mathbb{E}[\|\widehat{f}_\theta^t - f_\theta^t\|_\Theta^r] \leq \mathbb{E}[\|X_0\|^r] \Big(\sum_{j \geq t} \alpha_j(f) \Big)^r \quad \text{for all } t \in \mathbb{N}^*. \tag{3.1}$$

2. *If* **(A1(M))** *holds then $H_\theta^t \in \mathbb{L}^{r/2}(\mathcal{C}(\Theta, \mathcal{M}_m))$ and there exists $C > 0$ not depending on t such that*

$$\mathbb{E}[\|\widehat{H}_\theta^t - H_\theta^t\|_\Theta^{r/2}] \leq C \Big(\sum_{j \geq t} \alpha_j(M) \Big)^{r/2} \quad \text{for all } t \in \mathbb{N}^*. \tag{3.2}$$

3. *If* **(A1(H))** *holds then $H_\theta^t \in \mathbb{L}^{r/2}(\mathcal{C}(\Theta, \mathcal{M}_m))$ and*

$$\mathbb{E}[\|\widehat{H}_\theta^t - H_\theta^t\|_\Theta^{r/2}] \leq \mathbb{E}[\|X_0\|^r] \Big(\sum_{j \geq t} \alpha_j(H) \Big)^{r/2} \quad \text{for all } t \in \mathbb{N}^*. \tag{3.3}$$

Moreover, under any of the two last conditions and with **(A2)***, H_θ^t is an invertible matrix and $\left\| (\widehat{H}_\theta^t)^{-1} \right\|_\Theta \leq \underline{H}^{-1/m}$.*

The proof is given in Section 5.1.

3.2. *Strong consistency.* In the following theorem, we assume by convention that if **(A1(M))** holds then $\alpha_j(H) = 0$ and if **(A1(H))** holds then $\alpha_j(M) = 0$.

Theorem 1 *Assume that $\theta_0 \in \Theta$ for a compact subset $\Theta \subset \Theta(2)$. Let X be the stationary solution of the equation* (1.1)*. Let* **(A1(f))**, **(A2)** *and* **(Id)** *hold. Moreover, if* **(A1(M))** *or* **(A1(f))** *hold with*

$$\alpha_j(f) + \alpha_j(M) + \alpha_j(H) = O(j^{-\ell}) \quad \text{for some } \ \ell > 3/2, \tag{3.4}$$

then the QMLE $\widehat{\theta}_n$ defined by (1.5) *is strongly consistent, i.e. $\widehat{\theta}_n \xrightarrow[n\to\infty]{a.s.} \theta_0$.*

The proof is given in Section 5.3.

3.3. *Asymptotic normality.* We use the following convention: if **(A3(M))** holds then $\alpha_j^{(1)}(H) = 0$ and if **(A3(H))** holds then $\alpha_j^{(1)}(M) = 0$.

Theorem 2 *Assume that $\theta \in \overset{\circ}{\Theta}$ with $\overset{\circ}{\Theta} \subset \Theta(4)$ where $\overset{\circ}{\Theta}$ denotes the interior of a compact subset $\Theta \subset \mathbb{R}^d$. Let X be the stationary solution of the equation* (1.1). *Assume that the conditions of Theorem 1 and* **(A3(f))**, **(Var)** *hold. Moreover, if* **(A3(M))** *or* **(A3(H))** *holds with*

$$\alpha_j^{(1)}(f) + \alpha_j^{(1)}(M) + \alpha_j^{(1)}(H) = O(j^{-\ell'}) \quad \text{for some} \quad \ell' > 3/2, \tag{3.5}$$

then the QMLE $\widehat{\theta}_n$ is strongly consistent and asymptotically normal, i.e.,

$$\sqrt{n}(\widehat{\theta}_n - \theta_0) \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}_d(0\ ,\ F(\theta_0)^{-1}G(\theta_0)F(\theta_0)^{-1}), \tag{3.6}$$

where the matrices $F(\theta_0)$ and $G(\theta_0)$ are defined in (5.10) *and* (5.14) *respectively.*

The proof is given in section 5.4.

4. Examples. In this section, the previous asymptotic results are applied to several examples. For ARCH, GARCH, AR and GARCH-ARMA processes, the consistency and asymptotic normality have already been settled and we compare the different conditions from the literature with ours. For other examples, such as TARCH, multivariate ARCH and NLARCH processes, the consistency and the asymptotic normality of the QMLE are novel results. Examples satisfying the conditions of Corollary 1 are studied first.

4.1. ARCH(∞) *processes.* By Remark 2.1, the set $\widetilde{\Theta}(r)$ is well-adapted to that case

$$\widetilde{\Theta}(r) = \Big\{ \theta \in \mathbb{R}^d \ \Big/ \ \sum_{j=1}^{\infty} b_j(\theta) < \Big(\mathbb{E}\big[|\xi_0|^r\big]\Big)^{-2/r} \Big\}. \tag{4.1}$$

For $\theta_0 \in \widetilde{\Theta}(r)$, the existence of a stationary solution and of its r-th order moments is also settled in Giraitis *et al.* [16]. For an excellent survey about results and applications of ARCH models, we refer the reader to Giraitis *et al.* [18]. Here we formulate a version of Theorems 1 and 2 adapted to the context.

Proposition 2 *Let Θ be a compact subset of $\widetilde{\Theta}(2)$ and X the stationary solution of (2.3). Assume that $\inf_{\theta\in\Theta} b_0(\theta) > 0$ and that $\theta \mapsto b_j(\theta)$ be continuous functions satisfying*

$$\sup_{\theta\in\Theta} b_j(\theta) = O(j^{-\ell}) \quad \textit{for some} \quad \ell > 3/2.$$

1. *If* **(Id)** *holds then the QMLE $\widehat{\theta}_n$ is strongly consistent.*
2. *Assume that $\theta_0 \in \overset{\circ}{\Theta}$ with $\overset{\circ}{\Theta} \subset \widetilde{\Theta}(4)$, that ξ_0^2 has a non-degenerate distribution. Let the functions $\theta \mapsto b_j(\theta)$ be 2-times continuously differentiable on Θ for all $j \in \mathbb{N}$ satisfying for all $(k,k') \in \{1,\dots,d\}^2$,*

$$\sup_{\theta\in\Theta}\Big|\frac{\partial b_j(\theta)}{\partial\theta_k}\Big| = O(j^{-\ell'}) \ \textit{for some} \ \ell' > 3/2 \ \textit{and} \ \sum_{g\geq 1}\sup_{\theta\in\Theta}\Big|\frac{\partial^2 b_j(\theta)}{\partial\theta_k\partial\theta_{k'}}\Big| < \infty;$$

If there exists a finite subset $A \subset \mathbb{N}$ such that

$$\Big(\Big(\frac{\partial b_j(\theta_0)}{\partial\theta_k}\Big)_{j\in A}\Big)_{1\leq k\leq d} \ \textit{is linearly independent,} \tag{4.2}$$

then the QMLE $\widehat{\theta}_n$ is asymptotically normal, i.e., it satisfies (3.6).

For the asymptotic normality of the QMLE we use the condition in equation (4.2) coming from Robinson and Zaffaroni [25] that ensures both **(Id)** and **(Var)**. Let us compare the results of Proposition 2 with those of Theorems 1 and 2 in Robinson and Zaffaroni [25]. Those authors obtained the almost sure convergence of the QMLE under moments of order $r > 2$ (instead of $r = 2$ here) and a decreasing rate $j^{-\ell}$ with $\ell > 1$ (instead of $\ell > 3/2$ here) for the sequence $(\sup_{\theta\in\Theta}|b_j(\theta)|)_{j\geq 1}$. Concerning the asymptotic normality for $r = 4$, their conditions on both the first derivatives of $\theta \mapsto b_j(\theta)$ are the same as in Proposition 2. They required also conditions on the third derivatives (nothing like this here).

4.2. GARCH(q,q') *models.* The GARCH(q,q') models have been introduced by Engle [12]. Here X is the stationary solution of

$$X_t = \sigma_t\xi_t, \qquad \sigma_t^2 = c_0(\theta_0) + \sum_{j=1}^{q} c_j(\theta_0)X_{t-j}^2 + \sum_{j=1}^{q'} d_j(\theta_0)\sigma_{t-j}^2, \tag{4.3}$$

where $c_j(\theta)$ and $d_j(\theta)$ are non negative real numbers for all $\theta \in \Theta$. This model can be embedded in the class of ARCH(∞) models (see Giraitis *et al.* [18]), as one needs to set for all $z \in \mathbb{C}$

$$b_0(\theta) := \frac{c_0(\theta)}{1-\sum_{j=1}^{q'} d_j(\theta)} \quad \text{and} \quad \sum_{i=1}^{\infty} b_i(\theta) z^i := \frac{\sum_{i=1}^{q} c_i(\theta) z^i}{1-\sum_{i=1}^{q'} d_i(\theta) z^i}. \tag{4.4}$$

In the last formula, both the polynomials are supposed to be coprime. The results of Theorems 1 and 2 lead to the consistency and asymptotic normality of the QMLE in that case. However our conditions are not as sharp as those in Berkes and Horvath [1], Francq and Zakoïan [15] or Straumann and Mikosch [26]. As a consequence of the expression (4.4), the sequence $(\sup_{\theta\in\Theta} b_j(\theta))_j$ decreases exponentially fast to 0 and **A1(H)** holds automatically. From Corollary 1, if $\theta_0 \in \widetilde{\Theta}(r)$, where $\widetilde{\Theta}(r)$ is defined as in (4.1), the GARCH process has solutions of order r as for ARCH(∞). For instance, if $q = q' = 1$ and $\theta = (c_0, c_1, d_1)$, we achieve the optimal condition of existence of a second-order stationary solution as $\widetilde{\Theta}(2) = \{c_1, d_1$ such that $c_1 > 0$ and $c_1 + d_1 < 1\}$. In the property below, we use the identification condition of Francq and Zakoïan [15].

Proposition 3 *Assume that Θ is a compact subset of $\widetilde{\Theta}(2)$ defined in* (4.1) *and that X is the stationary solution to* (4.3)*. Assume that* $\inf_{\theta\in\Theta} c_0(\theta) > 0$, *that $\theta \mapsto c_j(\theta)$ and $\theta \mapsto d_j(\theta)$ are continuous functions injective on Θ for all j. If ξ_0 has a non degenerate distribution then the QMLE $\widehat{\theta}_n$ is strongly consistent.*
Moreover if $\theta_0 \in \overset{\circ}{\Theta}$ with $\overset{\circ}{\Theta} \subset \widetilde{\Theta}(4)$, the functions $\theta \mapsto c_j(\theta)$ and $\Theta \mapsto d_j(\theta)$ are 2 times continuously differentiable on Θ satisfying for all $(k,k') \in \{1,\dots,d\}^2$,

$$\sup_{\theta\in\Theta}\Big|\frac{\partial b_j(\theta)}{\partial\theta_k}\Big| = O(j^{-\ell'}) \ \ \text{for some} \ \ \ell' > 3/2 \ \text{and} \ \sum_{j\geq 1}\sup_{\theta\in\Theta}\Big|\frac{\partial^2 b_j(\theta)}{\partial\theta_k\partial\theta_{k'}}\Big| < \infty.$$

then the QMLE $\widehat{\theta}_n$ is also asymptotically normal.

Our assumptions induce the finiteness of second order moments of X. However Jeantheau [19] proves that moment conditions for models satisfying the Markov property are not needed for the consistency of the QMLE. In the case $\theta = (c_0, c_1, \dots, c_q, d_1, \dots, d_{q'})$ the Proposition 3 simplifies:

Proposition 4 *Assume that Θ is a compact subset of $\widetilde{\Theta}(2)$ and that X is the solution of* (4.3)*. Then, if ξ_0^2 has a non-degenerate distribution, the QMLE $\widehat{\theta}_n$ is strongly consistent. Moreover if $\theta_0 \in \overset{\circ}{\Theta}$ with $\overset{\circ}{\Theta} \subset \widetilde{\Theta}(4)$, then $\widehat{\theta}_n$ is also asymptotically normal.*

4.3. TARCH(∞) *models.* The process X is called Threshold ARCH(∞) if it satisfies the equations

$$X_t = \sigma_t \xi_t, \tag{4.5}$$

$$\sigma_t = b_0(\theta_0) + \sum_{j=1}^{\infty}\Big[b_j^+(\theta_0)\max(X_{t-j},0) - b_j^-(\theta_0)\min(X_{t-j},0)\Big],$$

where the parameters $b_0(\theta)$, $b_j^+(\theta)$ and $b_j^-(\theta)$ are assumed to be non negative real numbers. This class of processes is a generalization of the class of TGARCH(p,q) processes (introduced by Rabemananjara and Zakoïan [23]) and AGARCH(p,q) processes (introduced by Ding *et al.* [8]). Here

$$\Theta(r) = \left\{\theta \in \mathbb{R}^d \Big/ \sum_{j=1}^{\infty} \max\left(b_j^-(\theta), b_j^+(\theta)\right) \le \Big(\mathbb{E}\big[|\xi_0|^r\big]\Big)^{-1/r}\right\}$$

since $\alpha_j(M,\theta) = \max\left(b_j^-(\theta), b_j^+(\theta)\right)$. Consequently, we can settle for the first time the strong consistency and asymptotic normality of the QMLE for TARCH(∞) models:

Proposition 5 *Let Θ be a compact subset of $\Theta(2)$, X be the stationary solution to* (4.5) *and assume that* **(Id)** *holds. Assume that* $\inf_{\theta\in\Theta} b_0(\theta) > 0$ *and*

$$\sup_{\theta\in\Theta}\max\left(b_j^-(\theta), b_j^+(\theta)\right) = O(j^{-\ell}) \quad \textit{for some} \quad \ell > 3/2,$$

then the QMLE is strongly consistent.
Moreover if $\theta_0 \in \overset{\circ}{\Theta}$ with $\overset{\circ}{\Theta} \subset \widetilde{\Theta}(4)$, assume that the functions $\theta \mapsto b_0(\theta)$, $\theta \mapsto b_j^+(\theta)$ and $\theta \mapsto b_j^-(\theta)$ are 2 times continuously differentiable on Θ satisfying,

$$\sup_{\theta\in\Theta}\max\Big(\Big|\frac{\partial b_j^+(\theta)}{\partial\theta_k}\Big|, \Big|\frac{\partial b_-^+(\theta)}{\partial\theta_k}\Big|\Big) = O(j^{-\ell'}) \quad \textit{for some} \quad \ell' > 3/2 \quad \textit{and}$$

$$\sum_{j\ge1}\sup_{\theta\in\Theta}\max\Big(\Big|\frac{\partial^2 b_j^+(\theta)}{\partial\theta_k\partial\theta_{k'}}\Big|, \Big|\frac{\partial^2 b_j^-(\theta)}{\partial\theta_k\partial\theta_{k'}}\Big|\Big) < \infty \ \textit{for all } (k,k') \in \{1,\dots,d\}^2.$$

If **(Var)** *holds then the QMLE $\widehat{\theta}_n$ is also asymptotically normal.*

4.4. *Multivariate* ARCH(∞) *processes.* The multivariate ARCH(∞) processes are defined as solutions to equation (1.1) where

$$H_\theta(X_{t-1}, X_{t-2},\dots) := B_0(\theta) + \sum_{i=1}^{\infty} B_i(\theta)X_{t-i}X'_{t-i}B'_i(\theta). \tag{4.6}$$

Here $B_i(\theta)$ is a non-negative definite $d\times d$ matrice for all $\theta \in \mathbb{R}^d$. As $\alpha_j(M,\theta) = B_j(\theta)$ we have

$$\Theta(r) = \left\{\theta \in \mathbb{R}^d \Big/ \sum_{j=1}^{\infty} B_j(\theta) < \Big(\mathbb{E}\big[\|\xi_0\|^r\big]\Big)^{-1/r}\right\}. \tag{4.7}$$

Proposition 6 *Let Θ be a compact subset of $\Theta(2)$ and X be the stationary solution to (1.1) when relation (4.6) holds. Assume that $\inf_{\theta\in\Theta}\det B_0(\theta) > 0$, **(Id)** holds and*

$$\|B_j(\theta)\|_\theta = O(j^{-\ell}) \quad \textit{for some} \quad \ell > 3/2.$$

Then the QMLE is strongly consistent.
Moreover, if $\theta_0 \in \overset{\circ}{\Theta}$ with $\overset{\circ}{\Theta} \subset \widetilde{\Theta}(4)$, assume that the functions $\theta \mapsto B_j(\theta)$ are 2 times continuously differentiable on Θ satisfying for all $(k,k') \in \{1,\dots,d\}^2$,

$$\Big\|\frac{\partial B_j(\theta)}{\partial\theta_k}\Big\|_\Theta = O(j^{-\ell'}) \quad \textit{for some} \quad \ell' > 3/2 \quad \textit{and} \quad \sum_{j\geq 1}\Big\|\frac{\partial^2 B_j(\theta)}{\partial\theta_k\partial\theta_{k'}}\Big\|_\Theta < \infty.$$

*If **(Var)** holds, then the QMLE $\widehat{\theta}_n$ is also asymptotically normal.*

For the best of our knowledge, the asymptotic behavior of the QMLE for such models is studied here for the first time.

4.5. *Multivariate* GARCH(q,q') *models.* Multivariate GARCH(q,q') models refer classically to both VEC and BEKK models. We refer the reader to Section 4.8 for VEC models which are subcases of ARMA-GARCH models. BEKK processes are solutions of equation (4.6) or equivalently

$$vec(H_t) = vec(C_0) + \sum_{i=1}^{q} C_i(\theta_0)^* vec(X_{t-i}X'_{t-i}) + \sum_{i=1}^{p} D_i(\theta_0)^* vec(H_{t-i}),$$

where vec is the operator that stacks together the columns of a matrix. For any $p\times k$ matrix A: $A_i^* = \sum_{j=1}^k A_{i,j}\otimes A_{i,j}$ for $i=1,\dots,p$, where $\otimes$ denoting the Kronecker product. The multivariate ARCH(∞) representation holds with B_j satisfying

$$B_0^* := \Big(1 - \sum_{j=1}^{q'} D_j^*\Big)^{-1} \times C_0^* \quad \text{and} \tag{4.8}$$

$$\sum_{i=1}^{\infty} B_i^* Z^i := \Big(1 - \sum_{i=1}^{q'} D_i^* Z^i\Big)^{-1} \times \sum_{i=1}^{q} C_i^* Z^i \quad \text{for all} \quad Z \in \mathbb{C}^m.$$

In the last formula, both the polynomials are supposed to be coprime.

The natural choice $\theta = (C_0, C_1, \dots, C_q, D_1, \dots, D_{q'})$ implies that **(A1(M))** and **(Var)** are satisfied. Using the identification condition of Comte and Lieberman [7], Proposition 6 becomes more simple:

Proposition 7 *Let Θ be a compact subset of $\Theta(2)$ defined in (4.7) and X be the stationary solution to (1.1) when relations (4.6) and (4.8) hold. If $\inf_{\theta\in\Theta}\det C_0(\theta)>0$ and* **(Id)** *holds, then the strong consistency holds. Moreover if $\theta_0\in\overset{\circ}{\Theta}$ with $\overset{\circ}{\Theta}\subset\widetilde{\Theta}(4)$, then $\widehat{\theta}_n$ is also asymptotically normal.*

The asymptotic normality was settled before by Comte and Lieberman [7] for $r=8$. Our result needs just moment of order $r=4$.

4.6. *Multivariate* NLARCH(∞) *models.* Let $\big(B_j(\theta)\big)_{j\geq 1}$ be a sequence of $m\times d$-matrices and $B_0(\theta)$ be a vector of $\mathbb{R}^m$. The multivariate LARCH(∞) models introduced by Doukhan *et al.* [9] are extensions of the univariate LARCH(∞) models of Giraits *et al.* [17]. They are defined as the stationary solution of the equation:

$$X_t=\zeta_t\Big(B_0(\theta_0)+\sum_{j=1}^{\infty}B_j(\theta_0)X_{t-j}\Big).$$

Notice that the innovations $(\zeta_t)_{t\in\mathbb{Z}}$ are here random matrices. In this context, the QMLE is not a suitable estimator since $\inf_{\theta\in\Theta}\det\big(H_\theta(x)\big)=0$ except in very specific cases. However, Doukhan and Wintenberger [10] proposed a generalization of LARCH(∞) models, so-called NLARCH(∞) models defined by the equation

$$X_t \;=\; \zeta_t\Big(B_0(\theta)+\sum_{j=1}^{\infty}B_j(\theta,X_{t-j})\Big), \tag{4.9}$$

where now $B_j(\theta,.):\mathbb{R}^m\to\mathbb{R}^p$ are $b_j(\theta)$-Lipschitz functions. If the matrices of the innovations are concentrated on the diagonal, we rewrite (4.9) as

$$X_t=M_{\theta_0}(X_{t-1},X_{t-2},\ldots)\xi_t,$$

where $(\xi_t)_i=(\zeta_t)_{i,i}$ and $\big(M_\theta(X_{t-1},X_{t-2},\ldots)\big)_{ij}=\delta_{ij}\cdot\big(B_0(\theta)+\sum_{k=1}^{\infty}B_k(\theta,X_{t-k})\big)_i$. For instance, consider the multidimensional extension of the TARCH models as

$$(B_j(\theta,x))_k=\sum_{i=1}^{m}B^+_{j,k,i}(\theta)\max(x_{j,i},0)+B^-_{j,k,i}(\theta)\min(x_{j,i},0),$$

where $B^+_{j,k,i}$ and $B^-_{j,k,i}$ are non-negative real numbers. For NLARCH(∞) models we have

$$\Theta(r)=\left\{\theta\in\mathbb{R}^d\;\Big/\;\sum_{j=1}^{\infty}\mathrm{Lip}_x(B_j(\theta,x))<\Big(\mathbb{E}\big[\|\xi_0\|^r\big]\Big)^{-1/r}\right\}.$$

Proposition 8 *Let Θ be a compact subset of $\Theta(2)$, X be the stationary solution to* (4.9) *and assume that* **(Id)** *holds. Assume that for all $j \in \mathbb{N}$, the vectors $B_j(\theta, .) \in [0, \infty[^p$, $\inf_{\theta\in\Theta} \|B_0(\theta)\| > 0$ and for all $j \in \mathbb{N}^*$,*

$$\|Lip_x(B_j(\theta, x))\|_\Theta = O(j^{-\ell}) \quad \textit{for some} \quad \ell > 3/2,$$

then the strong consistency holds.
Moreover if $\theta_0 \in \overset{\circ}{\Theta}$ with $\overset{\circ}{\Theta} \subset \widetilde{\Theta}(4)$, assume that the functions $\theta \mapsto B_j(\theta, .)$ are 2 times continuously differentiable on Θ and satisfy

$$\Big\|Lip_x \frac{\partial B_j(\theta, x)}{\partial \theta_k}\Big\|_\theta = O(j^{-\ell'}) \quad \textit{with} \quad \ell' > 3/2$$
$$\textit{and} \sum_{j\geq 1} \Big\|Lip_x \frac{\partial^2 B_j(\theta, x)}{\partial \theta_k \partial \theta_{k'}}\Big\|_\Theta < \infty \textit{ for all } (k, k') \in \{1, \dots, d\}^2.$$

If **(Var)** *holds, the QMLE $\widehat{\theta}_n$ is also asymptotically normal.*

4.7. *Multivariate non-linear* AR(∞) *models.* Let us focus on one example where $M = I_m$ and $f \neq 0$. In this context, **(A1(M))** is always satisfied and the QMLE coincides with the least squares error estimator. Here, we restrict ourselves to the cases where

$$f_\theta(X_{t-1}, X_{t-2}, \dots) = A_0(\theta_0) + \sum_{i=1}^{\infty} A_i(\theta_0, X_{t-i}), \tag{4.10}$$

where $A_i(\theta, .)$ are Lipschitz functions with values in positive definite $d \times d$ matrices. Here $\Theta(r)$ neither depends on r nor on the distribution of ξ_0:

$$\Theta(r) = \left\{ \theta \in \mathbb{R}^d \Big/ \sum_{j=1}^{\infty} \mathrm{Lip}_x A_j(\theta, x) < 1 \right\}.$$

Proposition 9 *Let Θ be a compact subset of $\Theta(2)$, X be the stationary solution to* (1.1) *when $M = I_m$ and* (4.10) *holds. Under* **(Id)** *and if for $j \in \mathbb{N}^*$,*

$$\|Lip_x A_j\|_\Theta = O(j^{-\ell}) \quad \textit{for some} \quad \ell > 3/2,$$

then the strong consistency holds.
Moreover if $\theta_0 \in \overset{\circ}{\Theta}$ with $\overset{\circ}{\Theta} \subset \widetilde{\Theta}(4)$, $\widehat{\theta}_n$ is also asymptotically normal as soon as functions $\theta \mapsto A_j(\theta, x)$ are 2 times continuously differentiable on Θ for all $x \in \mathbb{R}^m$, **(Var)** *holds and*

$$\Big\| Lip_x \frac{\partial A_j}{\partial \theta_k} \Big\|_\Theta = O\big(j^{-\ell'}\big) \quad \textit{for some} \quad \ell' > 3/2 \textit{ and}$$
$$\sum_{j=1}^\infty \Big\| Lip_x \frac{\partial^2 A_j}{\partial \theta_k \theta_{k'}} \Big\|_\Theta < \infty \textit{ for all } k, k' \in \{1, \ldots, d\}.$$

4.8. *Multivariate* ARMA-GARCH *models.* Here M_θ is concentrated on its diagonal and f is not necessarily identically zero. If $f \equiv 0$, the model coincides with the VEC-GARCH model, see Jeantheau [19]. Multidimensional ARMA-GARCH processes were introduced by Ling and McAleer [21] as the solution of the system of equations

$$\left\{ \begin{array}{l} \Phi_\theta(L) \cdot X_t = \Psi_\theta(L) \cdot \varepsilon_t, \\ \varepsilon_t = M_\theta(X_{t-1}, X_{t-2}, \ldots)\xi_t, \end{array} \right. \tag{4.11}$$

with $\mathrm{diag}(H_\theta^t) = C_0(\theta) + \sum_{i=1}^q C_i(\theta)\mathrm{diag}(\varepsilon_{t-i}\varepsilon'_{t-i}) + \sum_{i=1}^{q'} D_i(\theta)\mathrm{diag}(H_\theta^{t-i})$. Here $C_0(\theta)$, $C_i(\theta)$ and $D_j(\theta)$ are positive definite matrices, $\mathrm{diag}A$ is the diagonal of the matrix A, $\Phi_\theta(L) = I_m - \Phi_1 L - \cdots - \Phi_s L^s$ and $\Psi_\theta(L) = I_m - \Psi_1 L - \cdots - \Psi_{s'} L^{s'}$ are polynomials in the lag operator L and Φ_i and Ψ_j are squared matrix. We define for all $\theta \in \mathbb{R}^d$

$$\begin{array}{rcl} \Gamma_\theta(L) & := & I_m + \displaystyle\sum_{i=1}^\infty \Gamma_i(\theta) L^i = \Psi_\theta^{-1}(L)\Phi_\theta(L) \\ \text{and } \displaystyle\sum_{i=1}^\infty B_i(\theta) Z^i & := & \Big(1 - \displaystyle\sum_{i=1}^{q'} D_i(\theta) Z^i\Big)^{-1} \times \displaystyle\sum_{i=1}^q C_i(\theta) Z^i \text{ for all } Z \in \mathbb{C}^m, \end{array}$$

where the polynomials of the right hand side are assumed to be coprime. The equation (4.11) has the representation (1.1) with $f_\theta(X_{t-1}, X_{t-2}, \ldots) := \sum_{i=1}^\infty \Gamma_i(\theta) X_{t-i}$. We can define,

$$\Theta(r) = \left\{ \theta \in \mathbb{R}^d \Big/ \sum_{i=1}^\infty \|\Gamma_i(\theta)\| + \Big(\mathbb{E}\big[\|\xi_0\|^r\big]\Big)^{1/r} \sum_{j=1}^\infty \|B_j(\theta)\| < 1 \right\}.$$

If $\theta_0 \in \Theta(r)$ then the existence of a solution is ensured. This existence condition is more explicit than the one of Theorem 2.1. of Ling and McAleer [21]. Now we give a version of Theorems 1 and 2 when

$$\theta = (\Phi_1, \ldots, \Phi_s, \Psi_1, \ldots, \Psi_{s'}, C_0, C_1, \ldots, C_q, D_1, \ldots, D_{q'}).$$

Proposition 10 *Let Θ be a compact subset of $\Theta(2)$, X be the stationary solution to the system* (4.11)*. If $\inf_{\theta\in\Theta} \det C_0(\theta) > 0$ and* **(Id)** *holds then $\widehat{\theta}_n$ is consistent.*
Moreover if $\theta_0 \in \overset{\circ}{\Theta}$ with $\overset{\circ}{\Theta} \subset \widetilde{\Theta}(4)$, $\widehat{\theta}_n$ is also asymptotically normal as soon as **(Var)** *holds.*

Ling and McAleer [21] also provided consistency and asymptotic normality of the QMLE. Proposition 10 improves their results. Notice that for VEC-GARCH models, Jeantheau [19] provided the consistency under a weaker condition.

5. Proofs. In this section the proofs of the main results are collected in the order of appearance in the paper. First we prove Proposition 1 and Corollary 1, then Lemma 1 that settles the invertibility of the QMLE. With the help of this property we prove the main theorems that state consistency and asymptotic normality of the QMLE.

5.1. *Proofs of Proposition 1 and Corollary 1.* We apply a result of Doukhan and Wintenberger [10] that gives conditions for the existence of a stationary solution of an equation of type

$$X_t = F(X_{t-1}, X_{t-2}, \ldots; \xi_t) \quad \text{a.e. for all} \quad t \in \mathbb{Z}. \tag{5.1}$$

If $\mathbb{E}\|\xi_0\|^r < \infty$ and F satisfies for $x = (x_i)_{i\geq 1}$, $y = (y_i)_{i\geq 1} \in (\mathbb{R}^m)^\infty$,

- $\mathbb{E}\|F(0;\xi_0)\|^r < \infty$;
- $\Big(\mathbb{E}\|F(x\,;\xi_0) - F(y\,;\xi_0)\|^r\Big)^{1/r} \leq \sum_{j\geq 1} a_j \|x_j - y_j\|$, with $\sum_{j\geq 1} a_j < 1$;

the existence of a unique causal stationary solution X of (5.1), such that $\mathbb{E}\big[\|X_0\|^r\big] < \infty$ is proved in [10]. We identify F from (1.1):

$$F(X_{t-1}, X_{t-2}, \ldots; \xi_t) = M_{\theta_0}(X_{t-1}, X_{t-2}, \ldots) \cdot \xi_t + f_{\theta_0}(X_{t-1}, X_{t-2}, \ldots).$$

Obviously, $\mathbb{E}\big[\|F(0;\xi_0)\|^r\big] < \infty$ if $\mathbb{E}\|\xi_0\|^r < \infty$ and we have

$$\begin{aligned}
&\Big(\mathbb{E}\|F(x\,;\xi_0) - F(y\,;\xi_0)\|^r\Big)^{1/r} \\
&\qquad\leq \Big(\mathbb{E}\|(M_{\theta_0}(x) - M_{\theta_0}(y)) \cdot \xi_0\|^r\Big)^{1/r} + \|f_{\theta_0}(x) - f_{\theta_0}(y)\| \\
&\qquad\leq \big(\mathbb{E}\|\xi_0\|^r\big)^{1/r} \|M_{\theta_0}(x) - M_{\theta_0}(y)\| + \|f_{\theta_0}(x) - f_{\theta_0}(y)\|.
\end{aligned}$$

The condition of Proposition 1 then implies those of [10] on F. In the context of Corollary 1, from $H_\theta(x) = \widetilde{H}_\theta(x^2)$ for all $x = (x_j)_{j\geq 1} \in \mathbb{R}^\infty$, we have

$$\begin{aligned}
\|\widetilde{H}_\theta(x) - \widetilde{H}_\theta(y)\| &\leq \sum_{j=1}^\infty \alpha_j(H,\theta)\|x_j - y_j\| \qquad \text{and,}\\
\left(\mathbb{E}\left[|M^2_{\theta_0}(x)\xi_0^2 - M^2_{\theta_0}(y)\xi_0^2|^{r/2}\right]\right)^{2/r} &= \left(\mathbb{E}\left[|\xi_0|^r\right]\right)^{2/r} |\widetilde{H}_{\theta_0}(x^2) - \widetilde{H}_{\theta_0}(y^2)|\\
&\leq \left(\mathbb{E}\left[|\xi_0|^r\right]\right)^{2/r} \sum_{j=1}^\infty \alpha^0_j(H,\theta_0)|x_j^2 - y_j^2|.
\end{aligned}$$

The results of [10] yield the existence in $\mathbb{L}^{r/2}$ of the solution $(X_t^2)_{t\in\mathbb{Z}}$ of the equation

$$X_t^2 = M^2_{\theta_0}(X_{t-1}, X_{t-2}, \ldots)\xi_t^2 = \widetilde{H}_{\theta_0}(X^2_{t-1}, X^2_{t-2}, \ldots)\xi_t^2 \qquad \text{a.e.}$$

Moreover, by [10] there exists a measurable function φ such that $X_t = \varphi(\xi_t, \xi_{t-1}, \ldots)$ for all $t \in \mathbb{Z}$. The ergodicity of X follows from the Proposition 4.3 in Krengel [20]; it states that if $(E, \mathcal{E})$ and $(\widetilde{E}, \widetilde{\mathcal{E}})$ are measurable spaces, $(v_t)_{t\in\mathbb{Z}}$ is a stationary ergodic sequence of E-valued random elements and $\varphi : (E^{\mathbb{N}}, \mathcal{E}^{\mathbb{N}}) \mapsto (\widetilde{E}, \widetilde{\mathcal{E}})$ is a measurable function then the sequence $(\widetilde{v}_t)_{t\in\mathbb{Z}}$ defined by $\widetilde{v}_t = \varphi(v_t, v_{t-1}, \ldots)$ is a stationary ergodic process.

5.2. *Proof of Lemma 1.* We treat the three assertions of the lemma one after the other.

1. Define $f_\theta^{t,p} = f_\theta(X_{t-1}, \ldots, X_{t-p}, 0, 0, \ldots)$ for all $t \in \mathbb{Z}$ and $p \in \mathbb{N}$. We have $f_\theta^{t,p} \in \mathbb{L}^r(\mathcal{C}(\Theta, \mathbb{R}^m))$ because $\theta_0 \in \Theta(r)$ and, using Corollary 1, all the following quantities are finite:

$$\begin{aligned}
\Big(\mathbb{E}\left[\|f_\theta^{t,p}\|^r_\Theta\right]\Big)^{1/r} &\leq \Big(\mathbb{E}\left[\|f_\theta^{t,0} - f_\theta^{t,p}\|^r_\Theta\right]\Big)^{1/r} + \Big(\mathbb{E}\left[\|f_\theta^{t,0}\|^r_\Theta\right]\Big)^{1/r}\\
&\leq \Big(\sum_{j\geq 1}\alpha_j(f)\Big)\Big(\mathbb{E}\left[\|X_0\|^r\right]\Big)^{1/r} + \|f_\theta(0)\|_\Theta.
\end{aligned}$$

For $p < q$

$$\mathbb{E}\left[\|f_\theta^{t,p} - f_\theta^{t,q}\|^r_\Theta\right] \leq \mathbb{E}\Big[\Big\|\sum_{p<j\leq q}\alpha_j(f)X_{t-j}\Big\|^r\Big] \leq \mathbb{E}\left[\|X_0\|^r\right]\Big(\sum_{p<j\leq q}\alpha_j(f)\Big)^r.$$

Since $\sum_{j\geq 1}\alpha_j(f) < \infty$, $(f_\theta^{t,p})_{p\geq 0}$ satisfies the Cauchy criteria in $\mathbb{L}^r(\mathcal{C}(\Theta, \mathbb{R}^m))$ and it converges to $f_\theta^{t,\infty}$, that is f_θ^t on $\sigma(X_{t_1}, \ldots, X_{t_n})$ for all $n \in \mathbb{N}^*$ and $t > t_1 > \cdots > t_n$ (those σ-algebras generate $\sigma(X_{t-1}, X_{t-2}, \ldots)$ and therefore

$f_\theta^{t,\infty} =_{a.s} f_\theta^t$).

2. Define $H_\theta^{t,p} = H_\theta(X_{t-1}, \ldots, X_{t-p}, 0, \ldots)$ for all $p \in \mathbb{N}$ and $t \in \mathbb{N}$. From Corollary 1, $\theta_0 \in \Theta(r)$ and common inequalities satisfied by matrix norms, $H_\theta^{t,p} \in \mathbb{L}^{r/2}(\mathcal{C}(\Theta, \mathcal{M}_m))$ since, denoting $M_\theta^{t,p} = M_\theta(X_{t-1}, \ldots, X_{t-p}, 0, \ldots)$,

$$\|H_\theta^{t,p}\|_\Theta^{r/2} \leq \|M_\theta^{t,p}\|_\Theta^r \leq \Big(\|M_\theta(0)\|_\Theta + \sum_{j=1}^\infty \|X_{t-j}\|\alpha_j(M)\Big)^r.$$

We conclude as above that $H_\theta^t \in \mathbb{L}^{r/2}(\mathcal{C}(\Theta, \mathcal{M}_m))$ by bounding, for $p < q$,

$$\|H_\theta^{t,p} - H_\theta^{t,q}\|_\Theta^{r/2} \leq \|M_\theta^{t,p} - M_\theta^{t,q}\|_\Theta^{r/2}\Big(\|M_\theta^{t,p}\|_\Theta^{r/2} + \|M_\theta^{t,q}\|_\Theta^{r/2}\Big).$$

The Cauchy-Schwarz inequality implies that

$$\begin{aligned}
\mathbb{E}[\|H_\theta^{t,p} - H_\theta^{t,q}\|_\Theta^{r/2}] &\leq \Big(\mathbb{E}[\|(M_\theta^{t,p} - M_\theta^{t,q}\|_\Theta^r]\Big)^{1/2} \\
&\quad \times\Big[\Big(\mathbb{E}[\|M_\theta^{t,p}\|_\Theta^r]\Big)^{1/2} + \Big(\mathbb{E}[\|M_\theta^{t,q}\|_\Theta^r]\Big)^{1/2}\Big] \\
&\leq B\Big(\mathbb{E}\Big[\Big(\sum_{p<j\leq q} \alpha_j(M)\|X_{t-j}\|\Big)^r\Big]\Big)^{1/2} \\
&\leq B\left(\mathbb{E}[\|X_0\|^r]\right)^{1/2}\Big(\sum_{p<j\leq q} \alpha_j(M)\Big)^{r/2}
\end{aligned}$$

for some constant $B > 0$.

3. First notice that $\|X_0 X_0'\| \leq \|X_0\|^2$. Next, as in the previous proofs, $(H_\theta^{t,p})_{p \in \mathbb{N}^*}$ converges to H_θ^t in $\mathbb{L}^{r/2}(\mathcal{C}(\Theta, \mathcal{M}_m))$. Thus there exists a subsequence $(p_k)_{k \in \mathbb{N}}$ such that $\|H_\theta^{t,p_k} - H_\theta^t\|_\Theta \xrightarrow[k\to\infty]{a.s.} 0$. Thanks to the continuity of the determinant, $(\det H_\theta^{t,p_k})_{k \in \mathbb{N}}$ also converges a.s. to $\det H_\theta^t$. Then $\det H_\theta^t \geq \underline{H}$, H_θ^t is an invertible matrix and in view of elementary relations between matrix norm and determinant $\left\|(\widehat{H}_\theta^t)^{-1}\right\|_\Theta \leq \underline{H}^{-1/m}$.

5.3. *Proof of Theorem 1.* The proof of the theorem is divided into two parts. In (i) a uniform (in θ) law of large numbers on $(\widehat{q}_t)_{t \in \mathbb{N}^*}$ (defined in (1.4)) is established. In (ii), it is proved that $L(\theta) := -\mathbb{E}(q_t(\theta))/2$ has a unique maximum in θ_0. Those two conditions lead to the consistency of $\widehat{\theta}_n$.

(i) Using Proposition 1, with $q_t = G(X_t, X_{t-1}, \cdots)$, one deduces that $(q_t)_{t \in \mathbb{Z}}$ (defined in (1.3)) is a stationary ergodic sequence. ¿From Straumann and Mikosch [26], we know that if $(v_t)_{t\in\mathbb{Z}}$ is a stationary ergodic sequence of random elements with values in $\mathbb{C}(\Theta,\mathbb{R}^m)$, then the uniform (in $\theta\in\Theta$) law of large numbers is implied by $\mathbb{E}\|v_0\|_\Theta<\infty$. As a consequence, $(q_t)_{t\in\mathbb{Z}}$ satisfies a uniform (in $\theta\in\Theta$) strong law of large numbers as soon as $\mathbb{E}\big[\sup_\theta|q_t(\theta)|\big]<\infty$. But, from the inequality $\log(x)\le x-1$ for all $x\in]0,\infty[$ and Lemma 1, for all $t\in\mathbb{Z}$,

$$|q_t(\theta)|\le\frac{\|X_t-f_t(\theta)\|^2}{(\underline{H})^{1/m}}+m\Big|\frac{1}{m}\log\underline{H}+\frac{\|H_\theta^t\|}{\underline{M}^{1/m}}-1\Big|\quad\text{for all }\theta\in\Theta$$

$$\Longrightarrow\quad\sup_{\theta\in\Theta}|q_t(\theta)|\le\frac{\|X_t-f_t(\theta)\|_\Theta^2}{(\underline{H})^{1/m}}+\Big|\log\underline{H}\Big|+m\times\frac{\|H_\theta^t\|_\Theta}{\underline{H}^{1/m}}.\tag{5.2}$$

But for all $t\in\mathbb{Z}$, $\mathbb{E}\|X_t\|^r<\infty$, see Corollary 1, and $\mathbb{E}\big[\|f_\theta^t\|_\Theta^r\big]+\mathbb{E}\big[\|H_\theta^t\|_\Theta^{r/2}\big]<\infty$, see Lemma 1. As a consequence, the right hand side of (5.2) has a finite first moment and therefore

$$\mathbb{E}\big[\sup_{\theta\in\Theta}|q_t(\theta)|\big]<\infty.$$

The uniform strong law of large numbers for $(q_t(\theta))$ directly follows and hence

$$\Big\|\frac{L_n(\theta)}{n}-L(\theta)\Big\|_\Theta\xrightarrow[n\to\infty]{a.s.}0\quad\text{with}\quad L(\theta):=-\frac{1}{2}\mathbb{E}\big[q_0(\theta)\big].\tag{5.3}$$

Now, one shows that $\dfrac{1}{n}\|\widehat{L}_n-L_n\|_\Theta\xrightarrow[n\to\infty]{a.s.}0$. Indeed, for all $\theta\in\Theta$ and $t\in\mathbb{N}^*$,

$$\begin{aligned}
&|\widehat{q}_t(\theta)-q_t(\theta)|\\
&=\Big(\log\det\widehat{H}_\theta^t-\log\det H_\theta^t\Big)+(X_t-\widehat{f}_\theta^t)'\big(\widehat{H}_\theta^t\big)^{-1}(X_t-\widehat{f}_\theta^t)\\
&\quad-(X_t-f_\theta^t)'\big(H_\theta^t\big)^{-1}(X_t-f_\theta^t)\\
&\quad\le|C|^{-1}\Big|\det(\widehat{H}_\theta^t)-\det(H_\theta^t)\Big|+(X_t-\widehat{f}_\theta^t)'\Big[\big(\widehat{H}_\theta^t\big)^{-1}-\big(H_\theta^t\big)^{-1}\Big](X_t-\widehat{f}_\theta^t)\\
&\quad+(2X_t-\widehat{f}_\theta^t-f_\theta^t)'\big(H_\theta^t\big)^{-1}(f_\theta^t-\widehat{f}_\theta^t)\\
&\le\ \underline{H}^{-1}\Big\|\det(\widehat{H}_\theta^t)-\det(H_\theta^t)\Big\|_\Theta+2(\|X_t\|+\|\widehat{f}_\theta^t\|_\Theta)\|\big(\widehat{H}_\theta^t\big)^{-1}-\big(H_\theta^t\big)^{-1}\|_\Theta\\
&\quad+\Big(2\|X_t\|+\|\widehat{f}_\theta^t\|_\Theta+\|f_\theta^t\|_\Theta\Big)\|\big(H_\theta^t\big)^{-1}\|_\Theta\|f_\theta^t-\widehat{f}_\theta^t\|_\Theta
\end{aligned}\tag{5.4}$$

by the mean value theorem, with $C\in[\det(H_\theta^t),\det(\widehat{H}_\theta^t)]$ and therefore $|C|>\underline{H}$. On the one hand,

$$\|\big(\widehat{H}_\theta^t\big)^{-1}-\big(H_\theta^t\big)^{-1}\|_\Theta\ \le\ \|\big(\widehat{H}_\theta^t\big)^{-1}\|_\Theta\|\widehat{H}_\theta^t-H_\theta^t\|_\Theta\cdot\|\big(H_\theta^t\big)^{-1}\|_\Theta.$$

On the other hand, for an invertible matrix $A \in \mathcal{M}_m(\mathbb{R})$, and $H \in \mathcal{M}_m(\mathbb{R})$,

$$\det(A+H) = \det(A) + \det(A) \cdot \mathrm{Tr}\big((A^{-1})'H\big) + o(\|H\|),$$

where $\big|\mathrm{Tr}\big((A^{-1})'H\big)\big| \leq \|A^{-1}\| \cdot \|H\|$. Using the relation $\big\|\big(H_\theta^t\big)^{-1}\big\|_\Theta \geq \underline{H}^{-m}$ for all $t \in \mathbb{Z}$, there exists $C > 0$ not depending on t such that inequality (5.4) becomes:

$$\sup_{\theta\in\Theta} |\widehat{q}_t(\theta) - q_t(\theta)| \leq C\big(\|X_t\| + \|\widehat{f}_\theta^t\|_\Theta + \|f_\theta^t\|_\Theta\big) \times \Big(\big\|\widehat{H}_\theta^t - H_\theta^t\big\|_\Theta + \|f_\theta^t - \widehat{f}_\theta^t\|_\Theta\Big)$$

¿From the Hölder and Minkowski inequalities and by virtue of $3/2 = 1+1/2$,

$$\begin{aligned}\mathbb{E}\big[\sup_{\theta\in\Theta} |\widehat{q}_t(\theta) - q_t(\theta)|^{2/3}\big] &\leq C\Big(\mathbb{E}\big[\|X_t\| + \|\widehat{f}_\theta^t\|_\Theta + \|f_\theta^t\|_\Theta\big]^2\Big)^{1/3} \\ &\quad \times \Big(\mathbb{E}\big[\|\widehat{H}_\theta^t - H_\theta^t\|_\Theta\big] + \mathbb{E}\big[\|f_\theta^t - \widehat{f}_\theta^t\|_\Theta\big]\Big)^{2/3} \\ &\leq C'\Big(\sum_{j\geq t} \big[\alpha_j(f) + \alpha_j(M)\big]\Big)^{2/3}, \qquad (5.5)\end{aligned}$$

with $C' > 0$ not depending on θ and t. Now, consider for $n \in \mathbb{N}^*$,

$$S_n := \sum_{t=1}^n \frac{1}{t} \sup_{\theta\in\Theta} \big|\,\widehat{q}_t(\theta) - q_t(\theta)\big|.$$

Applying the Kronecker lemma (see Feller [14], p. 238), if $\lim_{n\to\infty} S_n < \infty$ a.s. then $\frac{1}{n} \cdot \|\widehat{L}_n - L_n\|_\Theta \xrightarrow[n\to\infty]{a.s.} 0$. Following Feller's arguments, it remains to show that for all $\varepsilon > 0$,

$$\mathbb{P}(\forall n \in \mathbb{N},\ \ \exists m > n \text{ such that } |S_m - S_n| > \varepsilon) := \mathbb{P}(A) = 0.$$

Let $\varepsilon > 0$ and denote

$$A_{m,n} := \{|S_m - S_n| > \varepsilon\}$$

for $m > n$. Notice that $A = \bigcap_{n\in\mathbb{N}} \bigcup_{m>n} A_{m,n}$. For $n \in \mathbb{N}^*$, the sequence of sets $(A_{m,n})_{m>n}$ is obviously increasing, and if $A_n := \bigcup_{m>n} A_{m,n}$, then $\lim_{m\to\infty} \mathbb{P}(A_{m,n}) = \mathbb{P}(A_n)$. Observe that $(A_n)_{n\in\mathbb{N}}$ is a decreasing sequence of sets and thus,

$$\lim_{n\to\infty} \lim_{m\to\infty} \mathbb{P}(A_{m,n}) = \lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}(A).$$

It remains to bound $\mathbb{P}(A_{m,n})$. From the Bienaymé-Chebyshev inequality,

$$\begin{aligned}\mathbb{P}(A_{m,n}) &= \mathbb{P}\Big(\sum_{t=n+1}^{m}\frac{1}{t}\sup_{\theta\in\Theta}|\widehat{q}_t(\theta)-q_t(\theta)|>\varepsilon\Big)\\ &\leq \frac{1}{\varepsilon^{2/3}}\mathbb{E}\Big[\Big(\sum_{t=n+1}^{m}\frac{1}{t}\sup_{\theta\in\Theta}|\widehat{q}_t(\theta)-q_t(\theta)|\Big)^{2/3}\Big]\\ &\leq \frac{1}{\varepsilon^{2/3}}\sum_{t=n+1}^{m}\frac{1}{t^{2/3}}\mathbb{E}\Big[\sup_{\theta\in\Theta}|\widehat{q}_t(\theta)-q_t(\theta)|^{2/3}\Big].\end{aligned}$$

Using (5.5) and condition (3.4), since $\ell>3/2$, there exists $C>0$ such that

$$\Big(\sum_{j=t}^{\infty}\alpha_j(f)+\alpha_j(M)+\alpha_j(H)\Big)^{2/3}\leq\frac{C}{t^{2(\ell-1)/3}}.$$

Thus, $t^{-2/3}\mathbb{E}\Big[\sup_{\theta\in\Theta}|\widehat{q}_t(\theta)-q_t(\theta)|^{2/3}\Big]\leq C\big(t^{-2\ell/3}\big)$ for some $C>0$ and

$$\sum_{t=1}^{\infty}\frac{1}{t^{2/3}}\mathbb{E}\Big[\sup_{\theta\in\Theta}|\widehat{q}_t(\theta)-q_t(\theta)|^{2/3}\Big]<\infty \text{ as } \ell>3/2.$$

Thus $\lim_{n\to\infty}\lim_{m\to\infty}\mathbb{P}(A_{m,n}) \underset{n\to\infty}{\longrightarrow} 0$ and $\dfrac{1}{n}\cdot\|\widehat{L}_n-L_n\|_\Theta \xrightarrow[n\to\infty]{a.s.} 0$.

(ii) See Proposition 2.1. of Jantheau [19].

5.4. *Proof of Theorem 2.* Let V be a Banach space (thereafter $V=\mathbb{R}^m$ or $V=\mathcal{M}_m$) and $\mathcal{D}^{(2)}\mathcal{C}(\Theta,V)$ denote the Banch space of V-valued 2 times continuously differentiable functions on Θ equipped with the uniform norm

$$\|g\|_{(2),\Theta}=\|g\|_\Theta+\Big\|\frac{\partial g}{\partial\theta}\Big\|_\Theta+\Big\|\frac{\partial^2 g}{\partial\theta\partial\theta'}\Big\|_\Theta.$$

We start by proving the following preliminary lemma:

Lemma 2 *Let θ_0 belong to $\Theta(r)$ $(r\geq 2)$ and assume that* **(A3(f))** *and* **(A3(M))** *or* **(A3(H))** *hold. Then*

$$f_\theta^t\in\mathbb{L}^r(\mathcal{D}^{(2)}\mathcal{C}(\Theta,\mathbb{R}^m)) \text{ and } H_\theta^t\in\mathbb{L}^{r/2}(\mathcal{D}^{(2)}\mathcal{C}(\Theta,\mathcal{M}_m)).$$

In view of the results of Lemmas 1 and 2, the functions $\partial L_n(\theta)/\partial\theta$ and $\partial^2L_n(\theta)/\partial\theta^2$ are measurable and a.s. finite for all $\theta\in\Theta$. Their asymptotic properties are described in the next two lemmas

Lemma 3 *Let θ_0 belong to $\Theta(r)$ $(r \geq 4)$ and assume that* **(A3(f))** *and* **(A3(M))** *or* **(A3(H))** *hold, then*

$$n^{-1/2}\frac{\partial L_n(\theta_0)}{\partial\theta} \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}_d(0, G(\theta_0)), \tag{5.6}$$

where $G(\theta_0) = (G(\theta_0))_{1\leq i,j\leq d}$ is finite and its expression is given in (5.14).

Lemma 4 *Let θ_0 belong to $\Theta(r)$ $(r \geq 4)$ and assume that* **(A3(f))** *and* **(A3(M))** *or* **(A3(H))** *hold, then*

$$\Big\|\frac{1}{n}\frac{\partial^2 L_n(\theta)}{\partial\theta\partial\theta'} - \frac{\partial^2 L(\theta)}{\partial\theta\partial\theta'}\Big\|_\Theta \xrightarrow[n\to\infty]{a.s.} 0 \quad \text{with} \quad \frac{\partial^2 L(\theta)}{\partial\theta\partial\theta'} := -\frac{1}{2}\mathbb{E}\Big[\frac{\partial^2 q_0}{\partial\theta\partial\theta'}(\theta)\Big]. \tag{5.7}$$

We postponed the proofs of Lemmas 1-4 to the end of the Section and continue with the proof of Theorem 2. From Theorem 1, we have

$$\widehat{\theta}_n \xrightarrow[n\to\infty]{a.s.} \theta_0. \tag{5.8}$$

Since $\theta_0 \in \overset{\circ}{\Theta}$, a Taylor expansion of $\partial L_n(\theta_0)/\partial\theta_i \in \mathbb{R}$ implies

$$\frac{\partial L_n(\widehat{\theta}_n)}{\partial\theta_i} = \frac{\partial L_n(\theta_0)}{\partial\theta_i} + \frac{\partial^2 L_n(\overline{\theta}_{n,i})}{\partial\theta\partial\theta_i}(\widehat{\theta}_n - \theta_0), \tag{5.9}$$

for n sufficiently large such that the $\overline{\theta}_{n,i} \in \Theta$, which are between $\widehat{\theta}_n$ and θ_0 for all $1 \leq i \leq d$. Using equations (5.7) and (5.8), we conclude with the uniform convergence theorem that

$$F_n := -2\Big(\frac{1}{n}\frac{\partial^2 L_n(\overline{\theta}_{n,i})}{\partial\theta\partial\theta_i}\Big)_{1\leq i\leq d} \xrightarrow[n\to\infty]{a.s.} F(\theta_0).$$

One obtains $\big(F(\theta_0)\big)_{ij} = \mathbb{E}\Big[\partial^2 q_0(\theta_0)/\partial\theta_i\partial\theta_j\Big]$ for $1 \leq i,j \leq d$. With similar arguments as for (5.13), since $X_t - f^t_{\theta_0} = M_{\theta_0}\xi_t$, with ξ_t independent of $(X_{t-1}, X_{t-2}, \ldots)$,

$$\mathbb{E}\Big[(X_t - f^t_{\theta_0})'\frac{\partial^2 (H^t_\theta)^{-1}}{\partial\theta_i\partial\theta_j}(X_t - f^t_{\theta_0})\Big] =$$
$$2\mathbb{E}\Big[\mathrm{Tr}\Big((H^t_{\theta_0})^{-2}\frac{\partial H^t_{\theta_0}}{\partial\theta_j}\frac{\partial H^t_{\theta_0}}{\partial\theta_i}\Big) - \mathrm{Tr}\Big((H^t_{\theta_0})^{-1}\frac{\partial^2 H^t_{\theta_0}}{\partial\theta_j\partial\theta_i}\Big)\Big]$$

¿From equation (5.15), we then derive the explicit expression

$$\big(F(\theta_0)\big)_{ij} = \mathbb{E}\Big[2\Big(\frac{\partial f^t_{\theta_0}}{\partial\theta_j}\Big)'(H^t_{\theta_0})^{-1}\frac{\partial f^t_{\theta_0}}{\partial\theta_i} + \mathrm{Tr}\Big((H^t_{\theta_0})^{-2}\frac{\partial H^t_{\theta_0}}{\partial\theta_j}\frac{\partial H^t_{\theta_0}}{\partial\theta_i}\Big)\Big]. \tag{5.10}$$

Under Assumption **(Var)**, $F(\theta_0)$ is a positive definite $d \times d$ matrix. Indeed, for all $Y = (y_1, \ldots, y_d) \in \mathbb{R}^d$,

$$Y' F(\theta_0) Y = \mathbb{E}\Big[2\Big(\sum_{1\le i\le d} y_i \frac{\partial f^t_{\theta_0}}{\partial \theta_i}\Big)' \big(H^t_{\theta_0}\big)^{-1}\Big(\sum_{1\le i\le d} y_i \frac{\partial f^t_{\theta_0}}{\partial \theta_i}\Big) + \mathrm{Tr}\Big(\big(H^t_{\theta_0}\big)^{-2}\Big(\sum_{1\le i\le d} y_i \frac{\partial H^t_{\theta_0}}{\partial \theta_i}\Big)^2\Big)\Big].$$

These two terms are nonnegative and at least one of them is positive under Assumption **(Var)**. Then $F(\theta_0)$ is an invertible matrix and there exists n large enough such that F_n is an invertible matrix. Moreover, (5.9) implies,

$$n(\widehat{\theta}_n - \theta_0) = -2F_n^{-1}\Big(\frac{\partial L_n(\widehat{\theta}_n)}{\partial \theta} - \frac{\partial L_n(\theta_0)}{\partial \theta}\Big).$$

Therefore, if $\frac{1}{\sqrt{n}}\Big\|\frac{\partial L_n(\widehat{\theta}_n)}{\partial \theta}\Big\| \xrightarrow[n\to\infty]{\mathcal{P}} 0$, using Lemma 3 one obtains Theorem 2. Since $\frac{\partial \widehat{L}_n(\widehat{\theta}_n)}{\partial \theta} = 0$ ($\widehat{\theta}_n$ is a local extremum for $\widehat{L}_n$),

$$\mathbb{E}\Big[\frac{1}{\sqrt{n}}\Big\|\frac{\partial L_n}{\partial \theta} - \frac{\partial \widehat{L}_n}{\partial \theta}\Big\|_\Theta\Big] \xrightarrow[n\to\infty]{} 0. \tag{5.11}$$

Using the relation (5.12), the following inequality

$$|a_1 b_1 c_1 - a_2 b_2 c_2| \le |a_1 - a_2||b_2||c_2| + |a_1||b_1 - b_2||c_2| + |a_1||b_1||c_1 - c_2|$$

and the bounds $\|(\widehat{H}^t_\theta)^{-1}\|_\Theta \le \underline{H}^{-1/m}$, $\|(H^t_\theta)^{-1}\|_\Theta \le \underline{H}^{-1/m}$, one obtains:

$$\begin{aligned}
\Big\|\frac{\partial q_t(\theta)}{\partial \theta_i} - \frac{\partial \widehat{q}_t(\theta)}{\partial \theta_i}\Big\|_\Theta \le \frac{2}{\underline{H}^{1/m}}\Big[\Big\|\frac{\partial \widehat{f}^t_\theta}{\partial \theta_i} - \frac{\partial f^t_\theta}{\partial \theta_i}\Big\|_\Theta \|X_t - \widehat{f}^t_\theta\|_\Theta + \Big\|\frac{\partial f^t_\theta}{\partial \theta_i}\Big\|_\Theta \|\widehat{f}^t_\theta - f^t_\theta\|_\Theta\Big] \\
+2\Big\|\frac{\partial f^t_\theta}{\partial \theta_i}\Big\|_\Theta \Big\|(H^t_\theta)^{-1} - (\widehat{H}^t_\theta)^{-1}\Big\|_\Theta \|X_t - \widehat{f}^t_\theta\|_\Theta + \|\widehat{f}^t_\theta - f^t_\theta\|_\Theta \Big\|\frac{\partial (\widehat{H}^t_\theta)^{-1}}{\partial \theta_i}\Big\|_\Theta \|X_t - \widehat{f}^t_\theta\|_\Theta \\
+\|X - f^t_\theta\|_\Theta \|X_t - \widehat{f}^t_\theta\|_\Theta \Big\|\frac{\partial (H^t_\theta)^{-1}}{\partial \theta_i} - \frac{\partial (\widehat{H}^t_\theta)^{-1}}{\partial \theta_i}\Big\|_\Theta + \Big\|(\widehat{H}^t_\theta)^{-1}\Big\|_\Theta \Big\|\frac{\partial H^t_\theta}{\partial \theta_i} - \frac{\partial \widehat{H}^t_\theta}{\partial \theta_i}\Big\|_\Theta \\
+\Big\|(H^t_\theta)^{-1} - (\widehat{H}^t_\theta)\Big\|_\Theta \Big\|\frac{\partial (H^t_\theta)^{-1}}{\partial \theta_i}\Big\|_\Theta
\end{aligned}$$

Under **(A3(f))** and **(A3(M))** or **(A3(H))**, there exists $C > 0$ such that

$$\mathbb{E}\|f^t_\theta - \widehat{f}^t_\theta\|^r_\Theta \le C\Big(\sum_{j\ge t} \alpha_j(f)\Big)^r \text{and } \mathbb{E}\Big\|\frac{\partial f^t_\theta}{\partial \theta_i} - \frac{\partial \widehat{f}^t_\theta}{\partial \theta_i}\Big\|^r_\Theta \le C\Big(\sum_{j\ge t} \alpha^{(1)}_j(f)\Big)^r.$$

The differences $\mathbb{E}\|H_\theta^t - \widehat{H}_\theta^t\|_\Theta^{r/2} \le C\Big(\sum_{j\ge t}\alpha_j(M)\Big)^{r/2}$ can also be bounded:

$$\mathbb{E}\Big\|\frac{\partial H_\theta^t}{\partial\theta_i} - \frac{\partial \widehat{H}_\theta^t}{\partial\theta_i}\Big\|_\Theta^{r/2} \le C\Big(\Big(\sum_{j\ge t}\alpha_j(M)\Big)^{r/2} + \Big(\sum_{j\ge t}\alpha_j^{(1)}(M)\Big)^{r/2}\Big),$$
$$\mathbb{E}\Big\|\frac{\partial (H_\theta^t)^{-1}}{\partial\theta_i} - \frac{\partial (\widehat{H}_\theta^t)^{-1}}{\partial\theta_i}\Big\|_\Theta^{r/2} \le C\Big(\Big(\sum_{j\ge t}\alpha_j(M)\Big)^{r/2} + \Big(\sum_{j\ge t}\alpha_j^{(1)}(M)\Big)^{r/2}\Big).$$

Finally, using Hölder inequalities, it exists another constant $C \ge 0$ satisfying

$$\mathbb{E}\Big\|\frac{\partial q_t(\theta)}{\partial\theta_i} - \frac{\partial \widehat{q}_t(\theta)}{\partial\theta_i}\Big\|_\Theta \le C\sum_{j\ge t}\big(\alpha_j(f) + \alpha_j(M) + \alpha_j(H) + \alpha_j^{(1)}(f) + \alpha_j^{(1)}(M) + \alpha_j^{(1)}(H)\big).$$

Under (3.5), $\displaystyle\frac{1}{\sqrt{n}}\sum_{t=1}^n \mathbb{E}\Big\|\frac{\partial q_t(\theta)}{\partial\theta_i} - \frac{\partial \widehat{q}_t(\theta)}{\partial\theta_i}\Big\|_\Theta \underset{n\to\infty}{\longrightarrow} 0$, and Theorem 2 follows.

Proof of Lemma 2. Here, we focus on the case of H_θ under **(A3(f))** and **(A3(M))**. The other cases are simpler.

With the same method and notation as in the proof of Lemma 1, the result holds as soon as the function $\theta \in \Theta \to H_\theta^{t,p}$ is proved to satisfy a Cauchy criterion in $\mathbb{L}^{r/2}(\mathcal{D}^{(2)}\mathcal{C}(\Theta, \mathcal{M}_m))$. Using the proof of Lemma 1, we already have $\mathbb{E}\|H_\theta^{t,p}\|_\Theta^{r/2} < \infty$. It remains to bound the quantities

$$\mathbb{E}\Big\|\frac{\partial H_\theta^{t,p}}{\partial\theta_i}\Big\|_\Theta^{r/2} \text{ and } \mathbb{E}\Big\|\frac{\partial^2 H_\theta^{t,p}}{\partial\theta_i\partial\theta_j}\Big\|_\Theta^{r/2} \qquad \forall\, i,j \in \{1,\dots,d\}, \qquad \forall p \in \mathbb{N}^*.$$

Using Assumption **(A3(M))**:

$$\begin{aligned}\Big\|\frac{\partial H_\theta^{t,p}}{\partial\theta_i}\Big\|_\Theta &\le 2\|M_\theta^{t,p}\|_\Theta\Big\|\frac{\partial M_\theta^{t,p}}{\partial\theta_i}\Big\|_\Theta \\ &\le \Big(\|M_\theta(0)\|_\Theta + \sum_{j=1}^\infty \alpha_j(M)\|X_{t-j}\|\Big)\Big(\Big\|\frac{\partial M_\theta(0)}{\partial\theta_i}\Big\|_\Theta + \sum_{j=1}^\infty \alpha_j^{(1)}(M)\|X_{t-j}\|\Big).\end{aligned}$$

Using $\mathbb{E}\big[\|X_0\|^r\big] < \infty$ and the Hölder and Minkowsky inequalities:

$$\mathbb{E}\Big[\Big\|\frac{\partial H_\theta^{t,p}}{\partial\theta_i}\Big\|_\Theta^{r/2}\Big] \le C\Big(\|M_\theta(0)\|_\Theta^r + \mathbb{E}\big[\|X_0\|^r\big]\big(\sum_{j=1}^\infty \alpha_j(M)\big)^r\Big)^{1/2}$$

$$\times \Big(\Big\| \frac{\partial M_\theta(0)}{\partial \theta_i} \Big\|_\Theta^r + \mathbb{E}[\|X_0\|^r] \big(\sum_{j=1}^\infty \alpha_j^{(1)}(M) \big)^r \Big)^{1/2}.$$

In the same way, there exists another constant $C > 0$ such that

$$\begin{aligned} \mathbb{E} \Big\| \frac{\partial^2 H_\theta^{t,p}}{\partial \theta_i \partial \theta_j} \Big\|_\Theta^{r/2} \leq & C \Big[\Big(\big(\sum_{j=1}^\infty \alpha_j^{(1)}(M) \big)^r \big(\sum_{j=1}^\infty \alpha_j^{(1)}(M) \big)^r \Big)^{1/2} \\ & + \Big(\big(\sum_{j=1}^\infty \alpha_j(M) \big)^r \big(\sum_{j=1}^\infty \alpha_j^{(2)}(M) \big)^r \Big)^{1/2} \Big]. \end{aligned}$$

¿From $\sum_j \alpha_j(M) < \infty$, $\sum_j \alpha_j^{(1)}(M) < \infty$ and $\sum_j \alpha_j^{(2)}(M) < \infty$ we deduce that $\mathbb{E}[\|H_\theta^{t,p}\|_{(2),\Theta}^{r/2}] < \infty$ for all $p \in \mathbb{N}^*$. In the same way as in the proof of Lemma 1 we can also prove that the sequence $(H_\theta^{t,p})_{p \in \mathbb{N}^*}$ satisfies the Cauchy criterion in the Banach space $\mathbb{L}^{r/2}(\mathcal{D}^{(2)}\mathcal{C}(\Theta, \mathcal{M}_m))$. For the first derivatives, the result easily follows from the inequality

$$\Big\| \frac{\partial H_\theta^{t,p}}{\partial \theta_i} - \frac{\partial H_\theta^{t,q}}{\partial \theta_i} \Big\|_\Theta \leq 2 \|M_\theta^{t,p} - M_\theta^{t,q}\|_\Theta \Big\| \frac{\partial M_\theta^{t,p}}{\partial \theta_i} \Big\|_\Theta + 2 \|M_\theta^{t,q}\|_\Theta \Big\| \frac{\partial H_\theta^{t,p}}{\partial \theta_i} - \frac{\partial H_\theta^{t,q}}{\partial \theta_i} \Big\|_\Theta.$$

For the second derivatives, a similar argument finishes the proof.

Proof of Lemma 3. Simple calculations give the relations

$$\frac{\partial (H_\theta^t)^{-1}}{\partial \theta_k} = -(H_\theta^t)^{-1} \frac{\partial H_\theta^t}{\partial \theta_k} (H_\theta^t)^{-1} \text{ and } \frac{\partial \ln \det (H_\theta^t)}{\partial \theta_k} = \operatorname{Tr}\Big((H_\theta^t)^{-1} \frac{\partial H_\theta^t}{\partial \theta_k} \Big).$$

¿From Lemma 2, $\partial f_\theta^t / \partial \theta$, $\partial H_\theta^t / \partial \theta$ and $\big(\widehat{H}_\theta^t\big)^{-1}$ are a.s. finite. Then $\partial L_n(\theta)/\partial \theta$ is an a.s. finite measurable function satisfying, for all $1 \leq i \leq d$, $\partial L_n(\theta)/\partial \theta_i = -\frac{1}{2} \sum_{t=1}^n \partial q_t(\theta)/\partial \theta_i$ with

$$\begin{aligned} \frac{\partial q_t(\theta)}{\partial \theta_k} = & -2 \Big(\frac{\partial f_\theta^t}{\partial \theta_k} \Big)' (H_\theta^t)^{-1} (X_t - f_\theta^t) \\ & + (X_t - f_\theta^t)' \frac{\partial (H_\theta^t)^{-1}}{\partial \theta_k} (X_t - f_\theta^t) + \operatorname{Tr}\Big((H_\theta^t)^{-1} \frac{\partial H_\theta^t}{\partial \theta_k} \Big). \end{aligned} \tag{5.12}$$

Denoting $\mathcal{F}_t = \sigma(X_t, X_{t-1}, \ldots)$, let us prove that $\Big(\frac{\partial q_t(\theta_0)}{\partial \theta}, \mathcal{F}_t \Big)_{t \in \mathbb{Z}}$ is a $\mathbb{R}^m$-valued martingale difference process. Indeed, for all $t \in \mathbb{Z}$,

$$\mathbb{E}\big((X_t - f_{\theta_0}^t) | \mathcal{F}_t \big) = 0 \text{ and } \mathbb{E}\big((X_t - f_{\theta_0}^t)(X_t - f_{\theta_0}^t)' | \mathcal{F}_t \big) = H_{\theta_0}^t.$$

As a consequence,

$$\mathbb{E}\Big(\frac{\partial q_t(\theta_0)}{\partial\theta_k}|\mathcal{F}_t\Big)=\mathbb{E}\Big((X_t-f^t_{\theta_0})'\frac{\partial(H^t_{\theta_0})^{-1}}{\partial\theta_k}(X_t-f^t_{\theta_0})|\mathcal{F}_t\Big)+\mathrm{Tr}\Big((H^t_{\theta_0})^{-1}\frac{\partial H^t_{\theta_0}}{\partial\theta_k}\Big).$$

We conclude by noticing that the first term of the sum is equal to

$$\mathbb{E}\Big(\mathrm{Tr}\Big(\frac{\partial(H^t_{\theta_0})^{-1}}{\partial\theta_k}(X_t-f^t_{\theta_0})(X_t-f^t_{\theta_0})'\Big)|\mathcal{F}_t\Big)=\mathrm{Tr}\Big(\frac{\partial(H^t_{\theta_0})^{-1}}{\partial\theta_k}H^t_{\theta_0}\Big).$$

In order to apply the Central Limit Theorem for martingale-differences, see [3], we have to prove that $\mathbb{E}\Big[\Big\|\frac{\partial q_t(\theta_0)}{\partial\theta}\Big\|^2\Big]<\infty$. Using the relation $X_t-f^t_{\theta_0}=M^t_{\theta_0}\xi_t$ for all $t\in\mathbb{Z}$, then

$$\begin{aligned}\frac{\partial q_t(\theta_0)}{\partial\theta_k}=&-2\Big(\frac{\partial f^t_{\theta_0}}{\partial\theta_k}\Big)'(H^t_{\theta_0})^{-1}M^t_{\theta_0}\xi_t-\xi_t'M^t_{\theta_0}{}'(H^t_{\theta_0})^{-1\prime}\frac{\partial H^t_{\theta_0}}{\partial\theta_k}(H^t_{\theta_0})^{-1}M^t_{\theta_0}\xi_t\\&+\mathrm{Tr}\Big((H^t_{\theta_0})^{-1}\frac{\partial H^t_{\theta_0}}{\partial\theta_k}\Big)\end{aligned}$$

Let us compute the expectation of the square of the second term of the sum, with $\mathrm{Tr}(ABC)=\mathrm{Tr}(CAB)=\mathrm{Tr}(ACB)$ for symmetric matrices A, B and C,

$$\begin{aligned}&\mathbb{E}\Big[(\xi_t'\xi_t)^2\ \mathrm{Tr}\Big(M^t_{\theta_0}{}'(H^t_{\theta_0})^{-1}\frac{\partial H^t_{\theta_0}}{\partial\theta_k}(H^t_{\theta_0})^{-1}\frac{\partial H^t_{\theta_0}}{\partial\theta_k}(H^t_{\theta_0})^{-1}M^t_{\theta_0}\Big)\Big]\\&=\mathbb{E}\Big[(\xi_t'\xi_t)^2\ \mathrm{Tr}\Big((H^t_{\theta_0})^{-2}\Big(\frac{\partial H^t_{\theta_0}}{\partial\theta_k}\Big)^2\Big)\Big].\end{aligned}\tag{5.13}$$

Using this relation, the bound $\|(H^t_{\theta_0})^{-1}\|_\Theta\le\underline{H}^{-1/m}$ and the independence of ξ_t and $\mathcal{F}_t$, there exists $C>0$ such that

$$\begin{aligned}\mathbb{E}\Big[\Big(\frac{\partial q_t(\theta_0)}{\partial\theta_k}\Big)^2\Big]\le C&\Bigg(\mathbb{E}\Big[\Big\|\frac{\partial f^t_{\theta_0}}{\partial\theta_k}\Big\|^2\Big\|M^t_{\theta_0}\Big\|^2\Big]\times\mathbb{E}[\|\xi_t\|^2]+\mathbb{E}\Big[\Big\|\frac{\partial H^t_{\theta_0}}{\partial\theta_k}\Big\|^2\Big]\\&+\mathbb{E}[\|\xi_t'\xi_t\|^2]\times\mathbb{E}\Big[\Big\|\frac{\partial H^t_{\theta_0}}{\partial\theta_k}\Big\|^2\Big]\Bigg)\end{aligned}$$

Therefore, since $r\ge4$, the moment conditions for the CLT are fulfilled

$$\mathbb{E}\Big[\Big\|\frac{\partial q_t(\theta_0)}{\partial\theta}\Big\|^2\Big]=\sum_{k=1}^d\mathbb{E}\Big[\frac{\partial q_t(\theta_0)}{\partial\theta_k}\Big]^2<\infty.$$

We compute the asymptotic covariance matrix of $\dfrac{\partial q_t(\theta_0)}{\partial \theta}$. Thus, $(G(\theta_0))_{ij}$

$$
\begin{aligned}
&= \mathbb{E}\Big[\frac{\partial q_t(\theta_0)}{\partial \theta_i}\frac{\partial q_t(\theta_0)}{\partial \theta_j}\Big] \\
&= \mathbb{E}\Bigg[4\Big(\frac{\partial f^t_{\theta_0}}{\partial \theta_i}\Big)'\big(H^t_{\theta_0}\big)^{-1}\Big(\frac{\partial f^t_{\theta_0}}{\partial \theta_j}\Big) - \mathrm{Tr}\Big(\big(H^t_{\theta_0}\big)^{-1}\frac{\partial H^t_{\theta_0}}{\partial \theta_i}\Big)\mathrm{Tr}\Big(\big(H^t_{\theta_0}\big)^{-1}\frac{\partial H^t_{\theta_0}}{\partial \theta_j}\Big) \\
&\quad +p(m_4+(p-1))\,\mathrm{Tr}\Big(\big(H^t_{\theta_0}\big)^{-2}\frac{\partial H^t_{\theta_0}}{\partial \theta_i}\frac{\partial H^t_{\theta_0}}{\partial \theta_j}\Big)\Bigg]. \qquad (5.14)
\end{aligned}
$$

To simplify the expression, we assume here that ξ_t and $-\xi_t$ have the same distribution in order that $\mathbb{E}[\xi_t\xi_t' A\xi_t]=0$ for A a matrix.

Proof of the Lemma 4. ¿From the proof of Proposition 1 and from the result of Lemma 2, the second derivative process $(\partial^2 q_t(\theta)/\partial\theta^2)_{t\in\mathbb{Z}}$ is stationary ergodic (it is a measurable function of $X_t, X_{t-1},\ldots$). Therefore it satisfies a Uniform Law of Large Numbers (ULLN) if its first uniform moment is bounded.

¿From equation (5.12), the second partial derivatives of $q_t(\theta)$ are

$$
\begin{aligned}
\frac{\partial^2 q_t(\theta)}{\partial \theta_i \partial \theta_j} &= -2\Big(\frac{\partial^2 f^t_\theta}{\partial \theta_i\partial \theta_j}\Big)'\big(H^t_\theta\big)^{-1}(X_t-f^t_\theta) \\
&\quad +(X_t-f^t_\theta)'\frac{\partial^2 \big(H^t_\theta\big)^{-1}}{\partial \theta_i\partial \theta_j}(X_t-f^t_\theta) - 2\Big(\Big(\frac{\partial f^t_\theta}{\partial \theta_i}\Big)'\frac{\partial \big(H^t_\theta\big)^{-1}}{\partial \theta_j} \\
&\quad +\Big(\frac{\partial f^t_\theta}{\partial \theta_j}\Big)'\frac{\partial \big(H^t_\theta\big)^{-1}}{\partial \theta_i}\Big)(X_t-f^t_\theta) + 2\Big(\frac{\partial f^t_\theta}{\partial \theta_i}\Big)'\big(H^t_\theta\big)^{-1}\Big(\frac{\partial f^t_\theta}{\partial \theta_i}\Big) \\
&\quad +\mathrm{Tr}\Big(\Big(\frac{\partial \big(H^t_\theta\big)^{-1}}{\partial \theta_j}\Big)\Big(\frac{\partial H^t_\theta}{\partial \theta_i}\Big)\Big) + \mathrm{Tr}\Big(\big(H^t_\theta\big)^{-1}\Big(\frac{\partial^2 H^t_\theta}{\partial \theta_i\partial \theta_j}\Big)\Big). \qquad (5.15)
\end{aligned}
$$

Therefore, using the bound $\|\big(H^t_\theta\big)^{-1}\|_\Theta \le \underline{M}^{-1/m}$ of Lemma 1 and usual relations between norms and traces of matrix, there exists $C>0$ such that

$$
\begin{aligned}
\Big\|\frac{\partial^2 q_t(\theta)}{\partial \theta_i\partial \theta_j}\Big\|_\Theta &\le C\Big[\Big(\Big\|\frac{\partial^2 f^t_\theta}{\partial \theta_i\partial \theta_j}\Big\|_\Theta + \Big\|\frac{\partial H^t_\theta}{\partial \theta_j}\Big\|_\Theta\Big\|\frac{\partial f^t_\theta}{\partial \theta_i}\Big\|_\Theta + \Big\|\frac{\partial H^t_\theta}{\partial \theta_i}\Big\|_\Theta\Big\|\frac{\partial f^t_\theta}{\partial \theta_j}\Big\|\Big)\|X_t-f^t_\theta\|_\Theta \\
&\quad + \Big\|\frac{\partial^2 H^t_\theta}{\partial \theta_i\partial \theta_j}\Big\|_\Theta\|X_t-f^t_\theta\|^2_\Theta + \Big\|\frac{\partial f^t_\theta}{\partial \theta_i}\Big\|_\Theta\Big\|\frac{\partial f^t_\theta}{\partial \theta_j}\Big\|_\Theta + \Big\|\frac{\partial H^t_\theta}{\partial \theta_i}\Big\|_\Theta\Big\|\frac{\partial H^t_\theta}{\partial \theta_j}\Big\|_\Theta\Big].
\end{aligned}
$$

We conclude that $\mathbb{E}\Big\|\dfrac{\partial^2 q_t(\theta)}{\partial\theta_i\partial\theta_j}\Big\|_\Theta^{r/4} < \infty$ ($r \geq 4$) since, for $t \in \mathbb{Z}$, $1 \leq i, j \leq d$,

$$\mathbb{E}\big[\|X_t\|^r\big] < +\infty, \mathbb{E}\big[\|f_\theta^t\|_\Theta^r\big] < +\infty, \mathbb{E}\Big[\Big\|\frac{\partial f_\theta^t}{\partial\theta_i}\Big\|_\Theta^r\Big] < +\infty, \mathbb{E}\Big[\Big\|\frac{\partial^2 f_\theta^t}{\partial\theta_i\partial\theta_j}\Big\|_\Theta^r\Big] < +\infty;$$
$$\mathbb{E}\Big[\Big\|H_\theta^t\Big\|_\Theta^{r/2}\Big] < +\infty, \mathbb{E}\Big[\Big\|\frac{\partial H_\theta^t}{\partial\theta_i}\Big\|_\Theta^{r/2}\Big] < +\infty, \mathbb{E}\Big[\Big\|\frac{\partial^2 H_\theta^t}{\partial\theta_i\partial\theta_j}\Big\|_\Theta^{r/2}\Big] < \infty.$$

As a consequence, the ULLN holds for $\partial^2 q_t(\theta)/\partial\theta^2$.

Acknowledgements. We would are very grateful to Thomas Mikosch who made a critical review of the drafts and with whom we have worked on the final version of this paper.

REFERENCES

[1] Berkes, I., and Horváth, L. The efficiency of the estimators of the parameters in GARCH processes. *Ann. Statist. 32*, 2 (2004), 633–655.

[2] Berkes, I., Horváth, L., and Kokoszka, P. GARCH processes: structure and estimation. *Bernoulli 9* (2003), 201–227.

[3] Billingsley. *Convergence of Probability Measures.* John Wiley & Sons Inc., New York, 1968.

[4] Bollerslev, T. Modelling the coherence in short-run nominal exchange rates: a multivariate generalized ARCH approach. *Review of Economics and Statistics 72* (1990), 498–505.

[5] Bollerslev, T., and Wooldridge, J. M. Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances. *Econometric Rev. 11*, 2 (1992), 143–172.

[6] Bougerol, P. Kalman filtering with random coefficients and contractions. *Probab. Theory Related Fields 31* (1993), 942–959.

[7] Comte, F., and Lieberman, O. Asymptotic theory for multivariate GARCH processes. *J. Multivariate Anal. 84*, 1 (2003), 61–84.

[8] Ding, Z., Granger, C. W., and Engle, R. A long memory property of stock market returns and a new model. *J. Empirical Finance 1* (1993), 83–106.

[9] Doukhan, P., Teyssière, G., and Winant, P. A LARCH(∞) vector valued process. In *Dependence in Probability and Statistics* (New York, 2006), P. Bertail, P. Doukhan, and P. Soulier, Eds., Lectures Notes in Statistics, Springer, pp. 245–258.

[10] Doukhan, P., and Wintenberger, O. Weakly dependent chains with infinite memory. Preprint available on www.crest.fr/pageperso/doukhan/publications.htm.

[11] Dunsmuir, W., and Hannan, E. J. Vector linear time series models. *Adv. in Appl. Probability 8*, 2 (1976), 339–364.

[12] Engle, R. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica 50* (1982), 987–1007.

[13] Engle, R., and Kroner, K. Multivariate simultaneous generalized ARCH. *Econometric Theory 11* (1995), 122–150.

[14] Feller, W. *An Introduction to Probability Theory and its Applications*, vol. 2. Wiley, 1966.

[15] Francq, C., and Zakoïan, J.-M. Maximum likelihood estimation of pure garch and arma-garch processes. *Bernoulli 10* (2004), 605–637.

[16] Giraitis, L., Kokoszka, P., and Lepus, R. Stationary arch models: Dependence structure and central limit theorem. *Econometric Theory 16* (2000), 322.

[17] Giraitis, L., Leipus, R., Robinson, P. M., and Surgailis, D. LARCH, leverage and long memory. *Journal of Financial Econometrics 2*, 2 (2004), 177–210.

[18] Giraitis, L., Leipus, R., and Surgailis, D. Recent advances in ARCH modelling. In *Long-Memory in Economics* (2006), G. Teyssière and A. Kirman, Eds., Springer Verlag., pp. 3–38.

[19] Jeantheau, T. Strong consistency of estimators for multivariate arch models. *Econometric Theory 14*, 1 (1998), 70–86.

[20] Krengel, U. *Ergodic Theorems.* Gruyter Studies in Mathematics 6, Walter de Gruyter & Co., Berlin., 1985.

[21] Ling, S., and McAleer, M. Asymptotic theory for a vector ARMA-GARCH model. *Econometric Theory 19* (2003), 280–310.

[22] Mauricio, J. A. Exact maximum likelihood estimation of stationary vector ARMA models. *J. Amer. Statist. Assoc. 90* (1995), 282–291.

[23] Rabemananjara, R., and Zakoïan, J. Threshold ARCH models and asymmetries in volatility. *Journal of Applied Econometrics 8* (1993), 31–49.

[24] Robinson, P. Testing for strong serial correlation and dynamic conditional heteroscedasticity in multiple regression. *Journal of Econometrics 47* (1991), 67–84.

[25] Robinson, P., and Zaffaroni, P. Pseudo-maximum likelihood estimation of ARCH(∞) models. *Ann. Statist. 34* (2006), 1049–1074.

[26] Straumann, D., and Mikosch, T. Quasi-maximum-likelihood estimation in conditionally heteroscedastic time series: A stochastic recurrence equations approach. *Ann. Statist. 34*, 5 (2006), 24492495.

SAMOS
Centre Pierre Mendès-France
90 Rue de Tolbiac
75634 Paris Cedex 13,
E-mail: Jean-Marc.Bardet@univ-paris1.fr; Olivier.Wintenberger@univ-paris1.fr
URL: http://matisse.univ-paris1.fr/bardet/; http://wintenberger.fr